

## SUMMARIES

(without alignments),  
17041.707 Million cell updates/sec

```

Title:      US-10-02/-000-1
Perfect score: 2976
Sequence:   1 ttatagtcgcttgttaat.....aaaaaaaaaaaaaaaaaa 2976

```

Scoring table: IDENTITY-NUC  
Gapop 10.0 , Gapext 1.0

Searched: 2054640 seqs, 14551402878 residues

Maximum price: 45 summaries  
Listing first

Database :

```

1: GenEmb1:*
2: 1: qb.ba:*
3: 2: qb.htg:*
4: 3: qb.in:*
5: 4: qb.om:*
6: 5: qb.ov:*
7: 6: qb.pat:*
8: 7: qb.ph:*
9: 8: qb.pl:*
10: 9: qb.pr:*
11: 10: qb.ro:*
12: 11: qb.sts:*
13: 12: qb.sy:*
14: 13: qb.un:*
15: 14: qb.vi:*
16: 15: em.ba:*
17: 16: em.fun:*
18: 17: em.hum:*
19: 18: em.in:*
20: 19: em.mu:*
21: 20: em.om:*
22: 21: em.or:*
23: 22: em.ov:*
24: 23: em.pat:*
25: 24: em.ph:*
26: 25: em.pl:*
27: 26: em.ro:*
28: 27: em.sts:*
29: 28: em.un:*
30: 29: em.vi:*
31: 30: em.htg.hum:*
32: 31: em.htg.in:*
33: 32: em.htg.other:*
34: 33: em.htg.mus:*
35: 34: em.htg.pln:*
36: 35: em.htg.rod:*
37: 36: em.htg.mam:*
38: 37: em.htg.vrl:*
39: 38: em.sy:*
40: 39: em.htgo.hum:*
41: 40: em.htgo.mus:*
42: 41: em.htgo.other:*

```

4.1:  $\epsilon_{\text{unc}} = 3\%$   
Pred. No. is the number of results predicted by chance to have a

Result No.	Score	Query Match	Length	DB	ID	Description
1	427	14.3	4193	8	KFB6LUC	X05918 Kluyveromyces
2	406.2	13.2	38269	8	AF486581	AF486581 Uncultured
3	393.6	13.2	39305	8	SPU33009	U33009 Schizosacch
4	393.6	13.2	41799	8	SPBC1663	AL355920 Schizosac
5	393.6	13.2	85837	8	SPU33010	U33010 Schizosacch
6	392.6	13.2	2793	8	AF329731	AF329731 Volvarell
7	350.4	11.8	3578	1	AF005277	AF005277 Cellulomo
8	343.4	11.5	3710	1	ATUC06G1	M59852 A. tumefaci
9	305.2	10.3	2636	8	AB12933760	AJ2933760 Agaricus
10	270	9.1	110000	6	AE080928.2	Continuation (3 of
11	205.2	6.9	2601	6	AE489445	AX489445 Sequence
12	197.4	6.6	37445	1	AS89445	AL031013 Streptomy
13	184.4	6.2	3241	6	AR173232	AR173232 Sequence
14	184.4	6.2	17665	1	AF521337	AF521337 Streptomy
15	178.2	6.0	2760	1	SEY14337	Y14327 Saccharopol
16	175.2	5.9	12441	1	AF079762	AF079762 Streptomy
17	161	5.4	2256	6	E11369	E11369 DNA encodin
18	157.2	5.3	2640	1	TN6G1B	Z77856 T. neapolita
19	156.8	5.3	3179	1	CE06GB	D1068 C. gilvus ge
20	152.2	5.1	11763	1	AE005774	AE005774 Caulobact
21	149.4	4.8	14878	1	AE005849	AE005849 Caulobact
22	143.6	4.8	10615	1	AE012240	AE012240 Xanthomon
23	138.6	4.7	15052	1	AF055579	AF055579 Streptomy
24	138.6	4.7	50937	6	AR159871	AR159871 Sequence
25	137.4	4.6	21243	1	AE001690	AE001690 Thermotog
26	136.2	4.6	3265	1	RAU92808	U92808 Ruminococc
27	135.8	4.6	2166	6	AR205082	AR205082 Sequence
28	132.6	4.5	3158	1	RALBGLT	X15415 Ruminococcu
29	129.2	4.3	2075	1	ABGRMA	M1276 Clostridiace
30	127.6	4.3	2620	1	CSBG1Z	Z64064 S. asterocor
31	126.4	4.3	4537	1	AF090429	AF090429 Ascopirill
32	126.4	4.2	2346	1	PA1276021	AJ276021 Prevotell
33	126	4.2	12140	1	AE011777	AE011777 Xanthomon
34	123.2	4.1	110000	2	LMF1CJR36_-08	Continuation (9 of
35	118.2	4.0	3542	1	AB071166	Continuation (9 of
36	118.2	3.9	3955	1	MIBBOLA	AB071166 Glucose
37	114.4	3.8	1557	1	AE004598	L67134 Microdispor
38	111.6	3.8	15852	1	AE006929	AE006929 Pseudomon
39	111.6	3.8	36500	1	MFC18	AE004598 Mycobacter
40	109	3.7	13332	1	AE005788	AE005788 Caulobact
41	106.2	3.6	2370	1	PRN35425	U35425 Prevotella
42	104.4	3.5	2894	1	CE06GB	X15644 Clostridium
43	104.4	3.5	12882	1	AE013715	AE013715 Yersinia
44	104.4	3.5	150754	9	AE023491	AE023491 Homo sapi
45	104.4	3.5	258050	1	AJ414153	AJ414153 Yersinia

## ALIGNMENTS

	RESULT 1				
KBEGLOC			4193 bp	DNA linear	PLN 12-SEP-1993
LOCUS					
DEFINITION	KBEGLOC				
ACCESSION	Kluyveromyces fragilis beta-glucosidase gene.				
VERSION	X05918				
KEYWORDS	X05918.1 GI:2804				
SOURCE	beta-glucosidase.				
ORGANISM	Kluyveromyces marxianus.				
	Kluyveromyces marxiensis.				
	Eukaryota; Fungi; Ascomycota; Saccharomycetaceae; Kluyveromyces.				
	Saccharomycetales; Saccharomycetaceae; Kluyveromyces.				
REFERENCE	1 (bases 1 to 4193)				
AUTHORS	Raynal,A., Gerbaud,C., Francinques,M.C. and Guerineau,M.				
TITLE	Sequence and transcription of the beta-glucosidase gene of Kluyveromyces cerevisiae				
	Kluyveromyces fragilis cloned in Saccharomyces cerevisiae				

```

/strain="ATCC 12424"

```

```

repeat 1
307. .315

```

```

/note="TATA-Box 1; k=

```

16. .424

```

02. .483
note="not
214=

```

protein\_id="CAA29353.1"

MOGRPLGGRGFESFSEDPLYLAGM

ESSLI DA AKPA DA ENAG LIAK F...

EFGRVYGEDI E V G Y R Y E K I O F

Iteration	Repeat A"
1	.3445

```

1: 5001
2: Yeast formic acid
3: e="

```

Inverted repeat B" 3666

75 c	941 g	1239 t
------	-------	--------

```
0; Mismatches 7
```

1

Page 2

```

QY 1461 TACACCGTCGGCGCTA--CACCACCGTTCCTCCATTCCTAGGCGAGCAGTGCCTACGC 1518
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1752 TACACCGTAGGGCTTATTCACCAATTCGATTTGGAGTTGGCAGAGTAGTTGATC 1811
QY 1519 CCGACGGCGCTCCGGGATGCGCTGGAGGGTCTTCAACA--GCCCGCTGGTACCCCTTA 1576
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1812 GATCTCTCAAAACACAGAGATGCTGAAAATGCTGATTAATTCGCAAGTTTAACTCAAT 1871
QY 1577 CCGCAGCACA-----TTGAGAGCTCTCTCTCACCAGAGC 1613
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1872 CCGATTAACAGAGATCTGAAGATGAAGAACCACTTCACCTTACCAATTAATTC 1931
QY 1614 GACATGACCTGGTGGACTACTACACCCCGAGCGGAGACAG-----TGGTAC 1664
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1932 AATGTTACTTATTTGATTTCAAAACATGAGAAAGTGGATCCAAAGAACCTTACTTTT 1991
QY 1665 GCCGACATGAGAGGCGACATACCGCGGACGAGACTGCACCTACGAGCTGGCGCTGTC 1724
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1992 GTAACTTAACCGGACAGTACGTGCTCAAGAGATGATTAATCTTCACTTCA 2051
QY 1725 GTCGCGGCGACGGCAAGGGGTAGTACGACGCTGCTGCGACAGCCACCAAG 1784
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 2052 GTTATGCTTCTGTTGTTCTACTTAAGATGATGATTTATGACCAAAAGCAGAAC 2111
QY 1785 CAGGTCCCGCGGCGATGCTTCTGCGCTCCGACCCGCGAGAGAGCGGCGCGATCAT 1844
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 2112 CAGAAAGGGGTAGTTCTGCTTGGAGTGTGTACCAAAAGAACCAAAAGTTGACT 2171
QY 1845 CTCGTCAGGCGACACGATACGATTAAGATGATGCTGCGACCGACCTACAC 1904
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 2172 TTGAAGAGGGCGCAATTTATATGATTAAGATTAAGATGATGCTGCGACCTACAGT 2231
QY 1905 CTCAAGGCGACACCATATGCTCCCGCGACGCTCCCTCGCGTGGCGGCTGCAAGTCT 1964
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 2232 TTGTTGGGGA-----ATTCGGTGCAGAGTGGATTCAGCTGCTGATTAAGGCG 2282
QY 1965 ATGACGACACCGCGCAATATGAAAAGTCCCTCCCTCGCGACGAGGCGACGACGAGTCT 2024
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 2283 ATCGATGATGACGAGGATTAAGAACCGACCAAAATTAAGATTAAGATTAAGATTAAG 2342
QY 2025 ATCATCTCGCGGCGCTTACGCGGCTGAGGAGACCGAGGCGCGCGAGGAGATG 2084
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 2343 GTGTTGATTAATTTGATTAATTTGATTAATTTGATTAATTTGATTAATTTGATTAAT 2402
QY 2085 AAGCTCCCGGCGCTGAGCAGCTCATTTGCCAGCTGCGCGCGGAGAACCCAAACAC 2144
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 2403 GATTTGCGCAAAAGAAATGATTAATTTGCTGCTGTTTGAAGGCAAAATCAATACT 2462
QY 2145 GTCGTCGTCATACGAGGCGACCGCGGAGAGATGCTGCGTGCAGCGCCAGCGCC 2204
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 2463 GTTATGATTAACCAATCAAGGATCCCAATTTGATTTGCTGTTGATTAAGGAGCAATTCG 2522
QY 2205 GTATCTACGAGCTGTGAGCGGCGAGAGAGAGGCGCACTCCATTTGCCAGCTGCTTT 2264
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 2523 TAGTTCAAGCTTGTGATGAGGATGATTAATTTGATTAATTTGATTAATTTGATTAAT 2582
QY 2265 GCGGACATACACCTCGGCGAGCTGTCCTCAGCTTCCCAACGCGCTGAGGAGCAAC 2324
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 2583 GGTGACGTGCTTCAAAATGATTAATTTGCTGCTTGGCGCAATTAATTTGATTAAT 2642
QY 2325 CCGCGGTTCTCACTTCGCGACGAGCGGCGCGAGCTGTACGCGGAGAGCTTAC 2384
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 2643 CCAAGCTTTTAACTTCAAGACGAGTTGCGAGAGATTTGATTAAGGAGATTAATCTTT 2702
QY 2385 GTGCGGTACAGTACTACGAGTTTGGCGAGAGACATCAATTTCCCTTTGGCGACGCG 2444
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 2703 GTTGTGATTAAGTACTACCAAAAGCTTCAAAAGAGTAGCCCTTCCCTTGGATTAAT 2762
QY 2445 GTTCTCTACACACTTTTCCCTTCCATTTCTCGGTCTTCAAGAGAGCGCAAGCTG 2504
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 2763 CTATCTATACACATTCAGACTAGATATTTCTGACTTCAAGGTAACTGATTAAGATA 2822

```

```

QY 2505 AGCGTCCCTCTCCGTCGAGAGACACCGGCTCCGTCGCCGCGACAGTGCACGCTC 2564
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 2823 GATATTTAGTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2882
QY 2565 TACGTCAAGCCCTTCCAGCGGCGCAAGATTAACCGCCCCGTCAGAGAGCTCAAGGCTTC 2624
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 2883 GTCCTACTTACGGCTCTTAACCTTAAGCTTCGAGACCGGTTAAGAGCTTAAGGAGATTC 2942
QY 2625 GCAAGGTCGAGACTCAGCCCGCGGAGAGAGAGGCGGTGACAAATGAGAGAGAGAGAG 2684
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 2943 GAAAAAGTTCAATTTGAAACAGGAGTGAAGAGAGAGATTAATTTGATTAATTTGATA 3002
QY 2685 TACGTCCGCTGCTATTTGATGAGAGCGGAGATGCTGTGTGTGAAAAGGTCATAT 2744
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 3003 GATGCAATTTCTTACTTTTACAGAGAGAGCTGCTAATATGACGTTGAAGAGTGAATAC 3062
QY 2745 GAGGTATTCGT 2755
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 3063 TTGGTTTCACT 3073
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

RESULT 2
AF486581 38269 bp DNA linear BCT 21-MAR-2002
LOCUS Uncultured bacterium aminoglycoside 3'-N-acetyltransferase gene,
DEFINITION complete cds.
ACCESSION AF486581
VERSION AF486581.1 GI:19569565
KEYWORDS unclutred bacterium.
SOURCE unclutred bacterium.
ORGANISM unclutred bacterium
REFERENCE 1 (bases 1 to 38269)
AUTHORS Cappellano,C.M., Courtois,S., Ball,M., Francon,F.X., Normand,P.,
Heylinck,G., Guenineau,M., Martinez,A., Hopke,J., Kolyek,S.,
August,P.R., Osburne,M.S., Jeannin,P., Simonet,P. and
Pernodet,J.-L.
TITLE Recombinant Environmental Libraries Provide Access to Microbial
Diversity for Natural Products Drug Discovery
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 38269)
AUTHORS August,P.R. and Martinez,A.
TITLE Direct Submission
JOURNAL Submitted (21-FEB-2002) Molecular Biodiversity, Aventis Cambridge
Genomics Center, 26 Landsdowne Street, Cambridge, MA 02139, USA
FEATURES
source
1..38269
/organism="uncultured bacterium"
/db_xref="taxon:77133"
/note="Isolated from soil"
complement(30829..31617)
/note="confers resistance to kanamycin in Escherichia
coli; similar to known cognate proteins"
/codon_start=1
/transl_table=11
/product="aminoglycoside 3'-N-acetyltransferase"
/protein_id="AAL92107.1"
/db_xref="GI:19569566"
/transl_table="MTRSVATRSLSADLDSATGLADGAVLVYHALRQYKLYGEPDA
IIDLADVIGPAGTILGYCDMDLELRDDPSMRPIAIFEDRSRSTDNQIWPAL
RTPGALRSGSPGASMAALGAELEWFTAHADLYGYGDSPLKLEAGAKYLMGAP
LPMUTLHAHEADFPKRIIEYEAIVLDGFTVMFEFEFSPDGLPDEYFAT
IVAEFLATGKRGKRGVEASVYVPAAMVAWAFVDLEKMGKTL"
BASE COUNT 6318 a 12117 c 12998 g 6835 t 1 others
ORIGIN
Query Match 13.6%; Score 406.2; DB 1; Length 38269;
Best Local Similarity 50.8%; Pred. No. 4.8e-44;
Matches 1213; Conservative 0; Mismatches 1098; Indels 78; Gaps 7;
QY 355 TTGAGGCCATCTTGAAGAAGCTCACCCCTGCGGAGAGGTGATCTGCTGCTGATTCG 414
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 3224 TCGAGGAACTGTCGACCGGATGACGCTCGAGGAGAGGTGAGGCTGTATTCGGCGAGG 3203

```



Db 34416 CGAAGTGGTGCAGCTTACGCTGGCGCGGATGCG---ACGGTTCGCGCGCGCA 34472

QY 2608 AGCACTCAAGGGCTTCCCAAGGTCGAACGACACCCGGGAGACGAAAGCGGTGACAA 2667

Db 34473 AGCACTCAAGGGCTTCCCAAGGTCGAACGACACCCGGGAGACGAAAGCGGTGACAA 34532

QY 2668 TCGAGAGCAGAGAGTACGCTGCGCTATTTTGTATGAGAGCGGGA 2716

Db 34533 TCGAGCTGAGTACCGGGCTTCCGCTTACTACCGCAGGAGCGGAGCA 34581

RESULT 3  
SPU33009 39305 bp DNA linear PLN 14-NOV-1995

LOCUS Schizosaccharomyces pombe cosmid 1683.

DEFINITION U33009

ACCESSION U33009

VERSION U33009.1 GI:1063616

KEYWORDS

SOURCE Schizosaccharomyces pombe.

ORGANISM Schizosaccharomyces pombe

REFERENCE 1 (bases 1 to 39305)  
McCombie, W.R.  
Sequence analysis of a region of the fission yeast genome

AUTHORS

JOURNAL Unpublished

REFERENCE 2 (bases 1 to 39305)  
McCombie, W.R.  
Direct Submission

AUTHORS Submitted (02-AUG-1995) W. Richard McCombie, Cold Spring Harbor Lab, 1 Bungtown Rd, P.O. Box 100, Cold Spring Harbor, NY 11754, USA

JOURNAL Location/Qualifiers

FEATURES

source 1..39305  
/organism="Schizosaccharomyces pombe"  
/db\_xref="taxon:4896"  
/chromosome="2"  
/clone="cosmid 1683"  
/clone\_1bp="Mizukami, et al., 1993. Cell 73:121-132."  
BASE COUNT 12234 a 7052 c 7011 g 13008 t

ORIGIN

Query Match 13.28; Score 393.6; DB 8; Length 39305;  
Best Local Similarity 49.0%; Pred. No. 2.1e-42;  
Matches 1199; Conservative 0; Mismatches 1229; Indels 18; Gaps 5;

QY 339 ATGGCTGATATTGATGTTGAGGCGCATCTTGAGAAGCTCACCCGCGGAGAGTGCAT 338

Db 8497 ATGATGGAACATGATGTTGAAGATTATCAATCAATGATATAGTGAAGAGCTATG 8556

QY 399 CTGCTGCTGATATGACATCTCTGACACAAGAGCTCTCCCAAGCATGAGTCCCTCT 458

Db 8557 CTGCTTCCGCGACTGACCTTGTGACATCTGACACATCTCTCTTACATTTCTTCC 8616

QY 459 CTGCGCTTACAGTGGCCCAAGCGGCTGAAGAGGACCAAGTTCTCAATGGCGTCT 518

Db 8617 ATTGATTATTCAGACGCGCTTAATGATTCCTGGAAGTACGCTCTTCAATTTCTTCCA 8676

QY 519 GCGCGCTGCTTCCCTGCGGACGCTGCTGGTTCCACATTCACCAAACTGCTCGAA 578

Db 8677 TCCGCTTGTTCCTTCTGAGACAGCGCTAGGGGCTACTTTCGACAAAAGTTACTATTC 8736

QY 579 GAGCAGATAGATGATGAGCAAGAGGCGCATGCTGAAGAGTCCGATGATCTCGCG 638

Db 8737 GAAGTGTGTGAATTTAGAGAAAGCAAAAGCGAAGAGTGTAGTGTGTGTGGGT 8796

QY 639 CCGACTTCAACATGACAGCTCCCTCTCTGCTGAGAGTGGCTTCAATGATGATGAG 698

Db 8797 CCAACGGTAAATATCCATCTGAGACCTTAAATGATGAGGGTTTGAATATTTCTGAA 8856

QY 699 GATCCGTTCTGGGCGGCTTGGAGGCTGTATCCGCGGCGCATGAGACACTGGA 758

Db 8857 GATTCATTTATATAGGCTTTCAGACTAGTATGATTTCTTGATTCGAAAGTAAAAAC 8916

QY 759 GTGAGGCTATCATGACACTTTTGTGCAATGATAGAGAGACAGCCATGATGTTG 818

Db 8917 GTCCAACTATGATTAAGCAATTTTGTGTATGATATGAAAGTGAAGAACTCTGTC 8976

QY 819 CAGAGCATGCTCAAGGGGCGGCTCCGTAATATGACACACTCCGCTTCGATTTGCT 878

Db 8977 AGTTTACGCTCTCACAAGAGGCACTTGAAGAGTATATCTTATGCCATTTCAGTTGGCA 9036

QY 879 GTGAGACTCCAGCCGGGTGCTTCATGACGCGTACAAATGCAATGAGCTGTGTC 938

Db 9037 TGCATATATTCATTAATTCATATCAATCACTATGACTCTTCAACAAGGATGTTGAAC 9996

QY 939 TGCAGCAGAAACCTTAATATCTTGTATGATGATCTTTCGAAAGAAATGGGTTGGATGCG 998

Db 9097 GTTCCCAATCTGAAATTTATTAAGCAATTTTAAAGAAAGAGTGGGTAAGAGT 9156

QY 999 CTATCATGAGCAGCTGCTGCGGACATACATACAGACAGACCGCTTGGCGAGGCTTC 1058

Db 9157 ACTATATATCTGATTTGTTGGAATCTTATTCGTTGAGAAAGCTTATGATGAGGTTTG 9216

QY 1059 GACCTGAGATGCGCCGAGCTCCACGCTTCGAGAGG--AAACACTCAAGTTCAACGTC 1115

Db 9217 GATTTGAGATGCGCAGAAAGCGAGATTCGAAATGTCAACACTATTCACACTTAAGA 9276

QY 1116 TCCAGCGAAGCGCTTATTCACAGCTATGACCAAGGCTTGGAGATCTTCACTTC 1175

Db 9277 GGATCCAAAGGCTTTCGAGTCAATTTTATGAGAAAGCAAAAGATGCTGGAAGCTT 9336

QY 1176 GTCAAGAGTGTCTGCTCCGAGTACGAGAGAGAGAGGCGCCGAGCACTGTCACAAAC 1235

Db 9337 GTGAACATATCTCGGAGATATACAGAGCAAGAAATCATTTGATGATTAATATGATTC 9396

QY 1236 ACCCGCGAAGGCGAGCTCTCTCCGAAAGTTGGCAAGAGGCGATCGTGTCTGAAG 1295

Db 9397 TCTGTTTAAAGAGAGC---ACTTAAATAATTTTGAAGGACAGCTATGCTCTTGA 9453

QY 1296 AACGAGAACAGCTTCGCGCTTGCAGCAAGAAAGAGAGAGCGTATGTCGCGCCCAAC 1355

Db 9454 AATTAATAAGAACTTTTCCATTTGTCAGAGAGGCGACGTTGCACTTATGAGACCTAAC 9513

QY 1356 GCCAAGCAGGCAATATACAGAGGCGAGGCTTGCAGCACTGAGGCTTACTACGACATC 1415

Db 9514 GCAAGGTTTGTAAATTAACAGGAGGTGATGCGCAATTTGAAGCTTATATATCTG 9573

QY 1416 ACTCCCTTTCAGGCGCTCAGCAAGAGCTCGAGAGCGCCCATCTGACACGCTGGCGCC 1475

Db 9574 AGTATGATATGAGCATATACAGAGGAAATAGAGGATGTCAGAGATATCTCTTGGTGC 9633

QY 1476 TACAC-----ACGTTCTCCCATTTAGGGAGACATGGCTCAGCGCCAGCGCGCT 1529

Db 9634 CATTAATATCTAACTTCCAAATATTTGAATTTATGTAACCTTGAACCGGCAAG 9693

QY 1530 CCGGCGATCGCTGAGGCTTTCACAGAGCGCCCTGATACCCCTTAACCGCAGACATTT 1589

Db 9694 CATGATATGCTTTCGAAAGTTTATCTTGAACCTGCAACCTCCGAAATATGACTTTGATA 9753

QY 1590 GAGCAGCTCTTCTACCAAGAGAGGACATGACCTGTGTGACTACTACACCCCA---G 1646

Db 9754 GAGCATTAAGATTTAGAAAGATGAGTGTGCTTTTACGATTTATTTGAACCAAAATG 9813

QY 1647 GCGGAGACAGCTGTGATCCGACATGAGGCGACGTATACCGCGAGAGAGCTGCAC 1706

Db 9814 AAGGATGCTATTTCTATATATGACATTTGAAGTTATTTATTTCTGACAGAGACCTGTT 9873

QY 1707 TACGAGCTGCGCTGCTGCTGCGGACGCAAAAGGCGTACGTAGACAGCAGCTGTC 1766

Db 9874 TATGATTTGGGATATCTGTTGTTGAGAGGCTGCTGCTGTTATGATGATGATTTATTA 9933

QY 1767 GTGCAACAGCAGAGAGAGTCCCGGCGATGCTTCTTCTGCGCTCCGCAACCGCGAG 1826

Db 9934 ATCGACAAACAGCAAGCAACCCCTACGATCTACTTTGAAATTTGAAACATTAAGA 9993



```

misc-feature
join(3364..4004,4134..4872)
/gene="SPBC1683.01"
/Note="Match to PF00083 sugar_tr, Sugar (and other)
transporter score 132.74"
complement(5308..5478)
/Note="partial degraded TPI-1 type LTR"
5967..6986
/gene="SPBC1683.02"
5967..6986
/gene="SPBC1683.02"
/Note="SPBC1683.02, len:339, SIMILARITY: Saccharomyces
cerevisiae, ADA YEAST, probable adenosine deaminase, (347
aa), fasta scores: opt: 723, E():0, (37.6% identity in 351
aa)"
/codon_start=1
/label="SPBC1683.02
/protein_id="CAB91164.1"
/db_xref="GI:7801295"
/translation="MEDIKRETEKIPKALHILHETLEAEIKIKSHRNKIPKROSS
IEIKESYNHDLASTLEYEYEGVELLHEODFYDCTCYLKASQNYVAMFEDP
QIARRGISFEYIKGLIRARDMDPFIYISOLIMCFREMSFENAEETLASLPIK
SEIIGIGLSDENENNPRIKLFKFORAROLGYRLTCHDLOKNTTHIRLOLEDIG
ERIDGINIIDDELILKILERNIPTVCPSEMEIYVPGKAPETIRIMIDGLKVIN
SDPAYMHCFYITENFNLAOKASLTKEELVOICRNSFEAMVISEEKRHHYLEANSF
ASAYD"
6141..6956
/gene="SPBC1683.02"
/Note="Match to PF00962 A-deaminase, Adenosine/AMP
deaminase score 164.32"
complement(7448..9007)
/gene="SPBC1683.03c"
complement(7448..9007)
/gene="SPBC1683.03c"
/Note="SPBC1683.03c, len:519, SIMILARITY: Saccharomyces
cerevisiae, 008902, orf yor378w, (515 aa), fasta scores:
opt: 2052, E():0, (58.0% identity in 509 aa)"
/codon_start=1
/label="SPBC1683.03c
/product="WFS transporter of unknown specificity"
/protein_id="CAB91165.1"
/db_xref="GI:7801296"
/translation="MANNYKSLGTYTRNNRERESKDVONEDTPLPVOKISMLFIE
IIVFIACTAOLMTAGISGSIAPNNIGISGTNPBGOLSWFPAISYLVGFIILA
GRIGDIYGHKRMVLTGYTRFISLSGSFYAKSYMIPVCAIGIAPALIPAL
ALLGRVYPEKKNLITFALGAPAPNGFLGSVSGIFALSWMPWTITATVCFV
AIIGYFAIPHIEADEVEEKOFDYLAFEVSGVLINFSWNOGPAVWQVPTVYLL
IIGELSLVFLVLEKRYVOPILAPSMNSSEMGCVLICVAGMACFGIMVYIIMQLEN
LRFATPLLVAVOLTGVISGCAALITGYLKLKLPKIMVSMIAFTVGTILIAIAP
IHOTYMAOTFVSIIVPMGMDSPFPAATLMLSDPFRKHOGLIASIVSTVYNSISIG
LGIAGYETHLNHKGVDLIRGYSANWMTGTGPGGLGVCAILLVFASLYLKVGQKSPK
FPLIMKNTKAV"
11335..13833
/gene="SPBC1683.04"
11335..13833
/gene="SPBC1683.04"
/Note="SPBC1683.04, len:832, SIMILARITY: Kluyveromyces
marianus, BGUS_KLUMA, beta-glucosidase precursor, (845
aa), fasta scores: opt: 1274, E():0, (46.3% identity in
848 aa)"
/codon_start=1
/label="SPBC1683.04
/product="beta-glucosidase precursor (EC 3.2.1.21);
glycosyl hydrolase family 3"
/protein_id="CAB91166.1"
/db_xref="GI:7801297"
/translation="IMEDHVEDLITNOLISEKAMLSGTDLMTHTAIPRLNPSIRLS
DGPNGIRGTSPFNSPACPGATGATGATGLLEFGYGLAEAKAGSVYLGIT
VNIHGRPLNGRGEPSFSDSTLIGSIVYLLIGOSNVACIKHPCNMEDERNV
SIDYSORALREYVIMPOLACKYSNFKSMTSNKVKNGEHSORILLNIREKEMV
KGTIISDFGTYSIKKALIDAGLDEMGRKFRFRVNTIOLHYSKELESITLDERKN
VLKLVKSWMTAEHNCLEANDSSCLREALKFRASOIVLKLKKKLLPLSKRGTPA
VIGNAKVCNYSGGGSANLKRYYTSMWIDIAKIDVPEYALGCHNYMLPNIANLL

```

```

misc-feature
11428..13776
/gene="SPBC1683.04"
/Note="Match to PF00933 Glyco_hydro_3, Glycosyl hydrolase
family 3 score 687.70"
16074..16191
/Note="mRNA from spc06472 118 1"
17673..19352
/gene="SPBC1683.05"
17673..19352
/gene="SPBC1683.05"
/Note="SPBC1683.05, len:559, SIMILARITY: Saccharomyces
cerevisiae, THLY_YEAST, putative thiamine transporter
yor192c, (599 aa), fasta scores: opt: 828, E():0, (26.8%
identity in 512 aa)"
/codon_start=1
/label="SPBC1683.05
/product="NCS1 allantoinate transporter"
/protein_id="CAB91167.1"
/db_xref="GI:7801298"
/translation="MEDPKSDEKPDIGISEKNLDVFGFESSVDVYKGRFASFKTL
EISSPEKENDLRTPPPDRRHSADITLYMSNGISASAFRTGTSWENGSLPKOA
LIALIAGNVFIAMPPTLNGLSHSHIIPAVOSRASFETNTLILIRFAGLFGYC
TNVYAGACVOITILAIKFSRFSYKNNRIPADAGITSDLSIFYTWISFPHLRPE
YLORFPLIKSISTYACFAMLIFFLCNYSHYVMDQPFVSGRSWVKCALNSVA
GSTLAVVNDFTRYVKKHPTVYOMLLPLVAAVASAIGIVGASVKIMGTAMD
LOIANNVNSRGRRAAFPMGLIYVSMIAONISDVAANDLIEFFRYADIRAOV
IYITIGANAYVWKTLONGTAFIAPISLISFISPAAGIIPADKRNHKKYDIDFVN
PBGIRYNLGNMRALAFIACAPPLIPGAMNSINPSTMPDVIHLYYGYFSFM
TAFIYVGNLIVPEAKETLLEAVVPPKSNMELVDPSTLSGDKRWYITDY"
complement(20737..21669)
/gene="SPBC1683.06c"
complement(20737..21669)
/gene="SPBC1683.06c"
/Note="SPBC1683.06c, len:310, SIMILARITY: Escherichia
coli, YBEK_ECOLI, hypothetical 33.8 kd protein in
laus-911 intergenic region, (311 aa), fasta scores: opt:
838, E():0, (44.5% identity in 308 aa)"
/codon_start=1
/label="SPBC1683.06c
/product="putative inosine-uridine preferring nucleoside
hydrolase"

```

```

Query Match 13.2% Score 393.6; DB 8; Length 41799;
Best Local Similarity 49.0%; Pred. No.2,1e-42;
Matches 1199; Conservative 0; Mismatches 1229; Indels 18; Gaps 5;

```



QY 639 CCGACTATCAACATGACGCTCCCTCTCGGTGACGCTGCTGAGTGTGAG 698  
11635 CCACAGGTAAATATCATGCTGAGACCTTAATGAGGAGGTTTGAATCATTTCTGAA 11694  
QY 699 GATTCCTTCTGGCGGGCTGGAGAGCTGGGCTCTCATCCGGGCAATTCAGACCTGGA 758  
11695 GATTCATCTTTATCAGGCTTCTGACGCTAGTATGTTATTTCTTGAAATGCAATTA 11754  
QY 759 GTGACAGCTACGATCAGCACTTTTGTGCAATGATCAGAGGAGGAGGATGATG 818  
11755 GTCCAGGATGATGATTAAGATTTTGTGTATGATATGAGATGAGAAAGCACTCTG 11814  
QY 819 CAGAGCATGTGACGAGGAGGCTCTCGTGAATCTACGACTCCGTTCCAGATTCT 878  
11815 AGATGAGCTGCTACAAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 11874  
QY 879 GTGCGAGCTCCAGCGCGGCTGCTTCATGAGGAGGAGGAGGAGGAGGAGGAGG 938  
11875 TGCATATTTCAAAATTTCAATACCTTATGACTTCTTACACAGGATGATGACAC 11934  
QY 939 TCGACAGGAGACCTTAATATCTGATGAGGATGCTTCAGAAAGGATGAGGATG 998  
11935 GTTCCCATCTCGAATTTATTAACAATATTTAAGAAAGAGTGGAGAGTGGAAAGT 11994  
QY 999 CTAATCATGAGGAGCTGTGACGCACTACATACGATACCAAGAGGAGGAGGAG 1058  
11995 ACTATATATATGATGATGATGATGATGATGATGATGATGATGATGATGATG 12054  
QY 1059 GACCTGAGATGCTCCGAGCTCCAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1115  
12055 GATTTGAGATGCTCCAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 12114  
QY 1116 TCCAAAGGAGAGGCTTTATTCAGCTATGACAGAGGAGGAGGAGGAGGAGGAG 1175  
12115 GATTCAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 12174  
QY 1176 GTCAAGAGTGTGCTGCTCCGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1235  
12175 GTGAAACATCTCCGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 12334  
QY 1236 ACCCCGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1295  
12235 TCTTGTAAAGAGAGC---ACTTAAATAATTTGCAAGGAGGAGGAGGAGGAGG 12291  
QY 1236 AACGAGAACAGCTTCTGCTGAGCAAGAGGAGGAGGAGGAGGAGGAGGAGGAG 1355  
12292 AATAAAGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 12351  
QY 1356 GCCAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1415  
12352 GCAAAGGTTTGTAAATTTACAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 12411  
QY 1416 ACTCCCTTGAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1475  
12412 AGTATGATGATGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 12471  
QY 1476 TACAGC-----ACGCTTCTCCATCTTGAAGGAGGAGGAGGAGGAGGAGGAG 1539  
12472 CATACATATATCAACCTTCTCAATATGCAAAATTTATTAATGCAACCTTAGAAC 12531  
QY 1530 CCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1589  
12532 CATGATATGATGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 12591  
QY 1590 GACGAGCTTCTTACCAAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1646  
12592 GACGATATGATGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 12651  
QY 1647 GCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1706  
12652 AAGATGCTATTTCTATATGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 12711  
QY 1707 TACGAGCTGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1766  
12712 TATGATTTGGGATATCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 12771  
QY 1767 GTGCAACAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1826  
12772 ATGCAACAAAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 12831  
QY 1827 GAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1886  
12832 GAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 12891  
QY 1887 TCGCAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1946  
12892 AGTGCAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 12951  
QY 1947 GTGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 2006  
12952 ATTGGCTGTGTTAAAGTTATGATCTGAAACGAGATGATGATGATGATGATGATG 13011  
QY 2007 AAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 2066  
13012 AATGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 13071  
QY 2067 GCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 2126  
13072 GAGATGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 13131  
QY 2127 GCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 2186  
13132 CATTCATGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 13191  
QY 2187 CTCGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 2246  
13192 ATCTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 13251  
QY 2247 ATTGCGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 2306  
13252 TTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 13311  
QY 2307 AACGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 2366  
13312 AAAAACTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 13371  
QY 2367 TACGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 2426  
13372 TATGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 13431  
QY 2427 TTTCCCTTGGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 2486  
13432 TTTCCATTTGATGATGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 13491  
QY 2487 CACAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 2546  
13492 AATTTGATGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 13551  
QY 2547 GCGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 2606  
13552 GCAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 13608  
QY 2607 AAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 2666  
13609 AAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 13668  
QY 2667 ATGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 2726  
13669 ATTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 13728  
QY 2727 GTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 2772  
13729 TCTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 13774



LOCUS	SPU33010	85837 bp	DNA	Linear	PLN 14-NOV-1995
DEFINITION	Schizosaccharomyces pombe cosmid 359, 1198 and 1683.				
ACCESSION	U33010				
VERSION	U33010.1	GI:1063617			
KEYWORDS					
SOURCE					
ORGANISM	Schizosaccharomyces pombe. Schizosaccharomyces pombe. Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes; Schizosaccharomycetales; Schizosaccharomycetaceae; Schizosaccharomyces. 1 (bases 1 to 85837) McCombie, W.R. Sequence analysis of a region of the fission yeast genome				
REFERENCE					
AUTHORS	McCombie, W.R.				
TITLE	Sequence analysis of a region of the fission yeast genome				
JOURNAL	Unpublished				
AUTHORS	2 (bases 1 to 85837)				
TITLE	McCombie, W.R.				
JOURNAL	Direct Submission				
	Submitted (02-AUG-1995) W. Richard McCombie, Cold Spring Harbor Lab, 1 Bungtown Rd, P.O. Box 100, Cold Spring Harbor, NY 11754, USA				
FEATURES					
Source	1. 85837 /organism="Schizosaccharomyces pombe" /db_xref="taxon:4896" /chromosome="2" /clone="cosmid 359 (partial), 1198 and 1683" /clone_lib="Mizukami, et al., 1993. Cell 73:121-132." Location/Qualifiers				
BASE COUNT	27560 a 14781 c 15395 g 28101 t				
ORIGIN					
Query Match	13.2%; Score 393.6; DB 8; Length 85837;				
Best Local Similarity	49.0%; Pred. No. 1.9e-42;				
Matches 1199; Conservative	0; Mismatches 1229; Indels 18; Gaps 5;				
339	ATGGCTGATATTGATGTTAGAGCCATCTTGAGAAAGCTCACCTGGCCAGAAAGGTCAT	398			
DB	29116 ATGATGGAAACAATGATGTGAATTTGATCAATCAATTTGATTAAGTAATGAAGGCTATG	29175			
399	CTGCTGGCTGATGATGACTTCTGGACACAAAGGCTCTCCCAAGCATGAGATCCCTCT	458			
DB	29176 CTGCTTCCGGACACGACCTTTGGCATATGACAGCAATTCCTGCTTAAATTCCTTCC	29235			
459	CTCCGCTTACAGATGCCCCCAAGCGCTTAAGAGGACCAAGTTCTTCAATGGCGCCT	518			
DB	29236 ATTAGATTATCAGACGCGCCCTAATGATACCGTGAACATGCTTCTCAATCTTCACCA	29295			
519	GGGCGCTGTTCCCTCGCGGACAGCTGCTCGGTTCCACATTAACCAACTGTCGAA	578			
DB	29296 TCCGCTGTTTCCCTGCTGCGAGACAGCGCTAGGGGCTACTTTGACAAAAGTTACTATT	29355			
579	GAGGCAGTAAAGATGATGGGCAAGAGGCGCATGCTAAGATGCGCATGTATCTCGGC	638			
DB	29356 GAAGTTGGTGAATATTATACAGAAAGCAAAAGCGATTAGTGTGTTTGGAT	29415			
639	CCGACTATCAACATGCAAGCGTCCCTCGGAGGACGAGGCTTCAGTCGATTTGGTGAG	698			
DB	29416 CCAAGGTAATATTCATGCTGAGACCTTAATATGTAAGGGTTTGAATCATTTTCTGAA	29475			
699	GATCCCTTCTCGCGGCTTGGAGCGGCTCTCATCCGCGGCAATCAGACACTGGA	758			
DB	29476 GATTCTACTTTATCAGGTCTTGACAGTACTGTTATTTCTTGATTTGCAAAAGTAATAAC	29535			
759	GTCGACGCTACATCAAGCACTTTTGTGCAATGATCAGAGGACAGCGGCATGATGGAG	818			
DB	29536 GTTCAAGCATGCAATTAAGATTTTGTGTGAATGAATGAATGAATGAATGAATGAATGAAT	29595			
819	CAGAGCATGCTCAGGAGCGGCTCCCGGAATATTAAGCACTCCCGTTCAGATTGCT	878			
DB	29596 AGTATTGACGCTCTCACAAAGGCGACATTGAAGAGTATATCTTATGCAATTTCAAGTTGGCA	29655			
879	GTCGAGACTCCACGCGGGGTGCTTCAATGACGGCGCTACATATGGCATCAATGGCGGTGCG	938			

Db	29656	TGCAATATTTCAAAATTTCAATTCACCTTATGACTTCTTACAACAAGGTGTAATGGTAACAC	29715
Qy	939	TGCACGCGAAMCCTAAATATCTTGATGAGGATGCTTGGAAAGAAATGGGCTTGGGATGCC	998
Db	29716	GTTTTCCCATCTCGAATTTTTTATAGACAAATTTTTTAAGAAAGGTGGAGTGAAGAAGT	29775
Qy	999	CTAATCATGAGCAGCATGCTGATGCGACATACATACAGTACACAGAAAGCGTGTGGCAGGCGTC	1058
Db	29776	ACTATATATCTGATGATGGTTTGGAACTTATTCGTTGAAGAAGCTATAGATGCAGGTTTG	29835
Qy	1059	GACCTGAGATGCCCGGACCTCCACGCTTCCGAGAG---AACACTCAATTTCAACGTC	1115
Db	29836	GATTTGGAGATGCCAGGAACCCGAGATTTCGAAATGTCAACACATATTCAACACTTTAGTA	29895
Qy	1116	TCCAAAGGAAGCCCTTATTCACAGCTCATFTACACAGAGGGCTAGGAAAGTCTCACTTC	1175
Db	29896	GGATCCAAAGACCTTTGCGAGTCAATTTTAAATGAAAGCAAAAGATGTCTCAACCTT	29955
Qy	1176	GTCAAGAAGTGTCTGCTCCGCGAGTGCAGGAAGACGGCCCCGAGACGACTGTCAACAC	1235
Db	29956	GTGAACAATCTCTGCGACAAATACAGAACCAAAAATCTTGTGTAGCTAAATATATGATTTCC	30015
Qy	1236	ACCCCGAAACGGCAGCTCTCTCTCCGGAAGTGTGGCAACGAGGGCACTGTCTCTGAAG	1295
Db	30016	TCTTGTTTAAGGAAGC---ACTTAAAAAATTTTGAACCCAGTCCATGTCTCTTGAAA	30072
Qy	1296	AACGGAACAAAGTCTGCTGCTTGAGCAAGAAGAAAGAAAGACGCTGATTTGTGGCCCCAC	1355
Db	30073	AATTAAGAAACTTTTGCCATTTGTCAAGGAAGGCACTTTGCAGTATATGACCTAAC	30132
Qy	1356	GCCAAAGCAGCCACATACCAGCGCGGAGGCTCTGCCGACATCAGGCGCTACTACGAGTC	1415
Db	30133	GCAAAAGTTTGTAAATTATACAGCGAGGTGATCCGCAAAATTTGAACCTTATTTATACGTG	30192
Qy	1416	ACTCCCTTTGAGCGGCTCGCAAGAGAGTTCAGAGCGCCCATCGTACACGCTGCGCGCC	1475
Db	30193	AGTATGTATGATGGCATACACACGAAATTAAGCGGTGTCAGATATATGCTTGTTGTC	30252
Qy	1476	TACAC-----ACGTTCTCTCCATTTCTAGCGAGCAGTGCCTACGCCCGAGCGCGCT	1529
Db	30253	CATAACTATCAAAACCTTCCAAATATTTGCAAAATTTATTAGTCAACCTTAAAGACGGGCAAG	30312
Qy	1530	CGGGGATCGCGTGGAGGCTTTCACACGAGGCCCTGTGACCCCTTAACCGCAGACACTT	1589
Db	30313	CATGATACGTTGCGAAGTTTATCTTGAACTCGAACCTCGCAAAATTAAGAACTTGTATA	30372
Qy	1590	GACGAGCTTCTTCCACAAGACGACATGACCTGTGTGACTACTACCACCCCA---G	1646
Db	30373	GACGATTTACGATTTAAGAAATGAGATGTGCTGTTTTACGATTTATGTAACGACAAATVG	30432
Qy	1647	GCGGACAGACGAGTGAAGCCGACATGGAGGCGACGTACACCGCGCAGAGAGCTGCACC	1706
Db	30433	AAGGATGGCTATTTCTATATATGACATATGAAGTTATTATTAATTCCTGACAGSAGGCTT	30492
Qy	1707	TACGAGCTCGGCTCTGCTGTGCGGACGCGAAAGCGTACGTAGACGACCAGTCTGCT	1766
Db	30493	TATGAATTTGGGATATCTGTGTTTGGAAAGCGCTGCTGTATTATGATGATCTTTATTA	30552
Qy	1767	GTGCAACAGCCACCAAGAGGTCCCCCGGCGATGTCCTTTCGGTCCGCCACCGCGAG	1826
Db	30553	ATGCAACACAGCAAGCAAGCAAAACCCCTACGAATCATACTTTTGAATTTTGAACAATAGAA	30612
Qy	1827	GAGACGGGCGCATCATCTGTGCAAGGCAACAGCTACATCAAGTCAAGATCGATTTCGC	1886
Db	30613	GAGAAATTCATTTTATTTTAAAGAAAGGGAAGATATAAGTACAGGGTGCAGTATGGA	30672
Qy	1887	TCCGACCAACCTTACACCTCAAGGGCAGACACCATGTGTCCCGGCGAGGCTCCCTCGC	1946
Db	30673	AGTGCAGCCACTTACACGCTGTCACTAATCTGTCTCCAGTACGGGAAGAAATACTCT	30732
Qy	1947	GTCGGGGGCTCAAGTCTATTTAGCAGACAGCGCGAAATGAAAGCTCGCTCGCCCTCGCC	2006
Db	30733	ATTGGCTGTCTTAATTAATTTATGTACTCTGAAACGAGATAGATTACGACAGTGAAGGTTTGC	30792



QY 866 GTTCCAGATTGCTGTGCGAGACTCCAGCGGGGTGCTTATGAGCGGCTACAAATGGCAAT 925  
 Db 586 TTTTATTTTAACTCAAAAGTATGCTGCGCATATGATGAGCATCGCTTACCAACGAGT 645  
 QY 926 CAATGGGCTGCGGACGAGAACCCCTAAATATCTTATGATGAGTGTGGAAGAAATG 985  
 Db 646 GAACGGGCTTATATAGCGAGAACCCCTTCTGCTCAACAGAGTCTTACGTATATAGT 705  
 QY 986 GGGTGGGATGGCTTAATCATGAGCGAGTGTAGCGCATACAGTACACAGAACCGCT 1045  
 Db 706 GAATATCAAGGATTTGATGATGATGATGATGATGATGATGATGATGATGATGAT 765  
 QY 1046 TGTGGCAGGCTGACCTTCCAGATGCGCGGAGCTCCAGCTTCCAGAGT---GAACACT 1102  
 Db 766 CAAGCGCGGCTTACCTCGAATGCTGCTGATCATTAATATGGGTACCTGATGATGAT 825  
 QY 1103 CAATTTAAGCTTCAACAGGAAAGCCCTTATTCACGTATGACAGAGGCTTACGGA 1162  
 Db 826 CAACCGTATTTCAAGCTGCGAAGCTACCGCCAGACATCAAGAGACGAGAGT 885  
 QY 1163 AGTCTTCACTTGTGCAAGAGTGTCT-----GCTCCGAGTGTACAGAGAGGCTC 1216  
 Db 886 TGTATTTAGATTTAGTCAAGAGTGTCTCAAGAGAGGCTTCTCATGATGCGATG 945  
 QY 1217 CGAAGAGCTGTCAACAAACAGCCCGGAAAGCGAGCTCTCTCCGGAAGTGTGCAACGA 1276  
 Db 946 CGAGAGGCTTACCTCGAGCTTGAATCCGACAGCTCTATGCGGCGCATCGATGTGA 1005  
 QY 1277 GGGCATCTGCTGTGGAAGAGAGAACAGCTTCTGCTTGA-----GCAAGAA 1327  
 Db 1006 GTCCATCTCTTACTCAAAATATGATTAAGTCTCTCTTAACTCAAGAGAGCGGAT 1065  
 QY 1328 GAAAGAGAGCTGTATGTCGCGCCCAAGCGCAAGAGGCTTACAGAGGCGGCTC 1387  
 Db 1066 CAAGAGAGTCCCGCTGTGTGGCAATGCAAGAGCGCAAGTCTGTGCGGAGAGGCT 1125  
 QY 1388 TGCCGCACTAGGCGCTTACTAGCAGTACCTCTTGTGAGGCGCTTACAGAGGCTTGA 1447  
 Db 1126 TGGGCGCTGAGAGGCTTACTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCT 1185  
 QY 1448 GAGCGCGCATCTGTAACCC-----GTGCGCGCTTACACACAGCTTCTCTC 1492  
 Db 1186 ACCTATGCGCTAGAGCTTCACTTCACTTCACTTCACTTCACTTCACTTCACTTCA 1245  
 QY 1493 CATTTAGCGCAGAGTCTTCCAGCGCGGCGCTCCGCGGCTGAGAGGCTT 1552  
 Db 1246 CCTCAATGGGATCTGGAAGAGAGAGCGAGCGGAGAGTGTGAGAGTGTGAGAGTGT 1305  
 QY 1553 CAACGAGCGCGCTTGTACCC-----CTTACCGCGCAGACATTTGAGAGCTTCTT 1603  
 Db 1306 CCAGAGTCCGAGCAGACATGAGCGGCTTGTGACGAGCATTTCACTTCTTCTTCTT 1365  
 QY 1604 CACCAAGAGAGATGACAGCTGTGTGAGTACTTACACCCCAAGCGGCGAGACAGTGT 1663  
 Db 1366 TGACCAAAACAGAAATCTTATCATGACAGCTGTATCCCAAGGAAATTAACAAAGTGT 1425  
 QY 1664 GCGCAGATGAGAGG---CAGCTACACCGCGCAGAGAGAGTGTGAGCTTCTGAGCT 1720  
 Db 1426 GATGAGACTGAGAGGCTTACTTGAAGCAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1485  
 QY 1721 CGTGTCTGCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1780  
 Db 1486 TATTTCTGTGTGCGGCTTAACTCTGTGGTTGACGAGACATGATTAAGTGTGAG 1545  
 QY 1781 CAAGAGAGTCCCGCGAGATGCTTCTGAGTCCGCGCAGCGGAGAGAGAGAGAGAG 1840  
 Db 1546 CGCTCAACGCTGTGTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1605  
 QY 1841 CAATCTGCTCAAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1894  
 Db 1606 TCTCTCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1665  
 QY 1895 CACCTAACCTTCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1954

Db 1666 GGCTATGATGATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1725  
 QY 1955 CTGCAAGCTTATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2014  
 Db 1726 TGCAGAGGTTGCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1785  
 QY 2015 CGACAGGCTATCATCTGCGGCGCTTAAAGCGCAGTGTGAGAGAGAGAGAGAGAG 2074  
 Db 1786 TGTATGCTGATGATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1845  
 QY 2075 CGGAGAGATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2134  
 Db 1846 AAGGATTTGGGCTCTTCTGCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1905  
 QY 2135 CCCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2194  
 Db 1906 CTGGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1965  
 QY 2195 CACGCGCGCTATCATCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2254  
 Db 1966 TGTGCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2025  
 QY 2255 CGTGTCTTGTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2314  
 Db 2026 CGTGTCTTGTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2085  
 QY 2315 GCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2374  
 Db 2086 GAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2145  
 QY 2375 GAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2434  
 Db 2146 AGAGTGTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2205  
 QY 2435 TGCGCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2484  
 Db 2206 CGGCTTGTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2265  
 QY 2485 -----CTCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2533  
 Db 2266 CATTTATGATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2325  
 QY 2534 CTCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2593  
 Db 2226 TCCGATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2385  
 QY 2594 TAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2653  
 Db 2386 GAGCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2445  
 QY 2654 GAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2713  
 Db 2446 GAGCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2505  
 QY 2714 GAGATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2773  
 Db 2506 TCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2565  
 QY 2774 GAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2833  
 Db 2566 AAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2595

RESULT 7  
 AF005277 3578 bp DNA linear BCT 02-MAR-1998  
 LOCUS AF005277  
 DEFINITION Cellulomonas blatzotea cellulobiose (cba) gene, complete cds.  
 ACCESSION AF005277  
 VERSION AF005277.1 GI:2921739  
 KEYWORDS  
 SOURCE Cellulomonas blatzotea  
 ORGANISM Cellulomonas blatzotea



1824 GTGACAGTGGGCTCCAGCGGCTGCTGACGCGACCGCTGTGTGGCTGCTCCACCCG 1883  
 1872 AAGATGAGTTTCGCTCCGACCCACCTACACCTTGAAGGCGACACATCTGCTCCCGG 1931  
 1884 AGGTGACGCTCGGCGACCGCCGACCGGACCGGCGCGCGCGCGCGCGCGCTC 1943  
 1932 CACGGC-----TCCCTCCGCTGCGGCGCTGCAAGTCAATTACACACCGCGAA 1982  
 1944 GCGGGGCTCGGGCTGTGTGGGCTGCTGCGCGCGCGCGCGCGCGCGCGCGCTC 2003  
 1983 ATGAAAAGTCCGTCGCGCTCCGCAAGAGACACAGCTATCTATCTGCGGCGCTT 2042  
 2004 ATACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCTC 2063  
 2043 AACGCCATGAGAGACCGACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCTC 2102  
 2064 ACCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2123  
 2103 GACCACTCATTTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCTC 2162  
 2124 GAGCGCTGCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCTC 2183  
 2163 GGCACCCCGCGAGAGATGCTGCTGCGCGCGCGCGCGCGCGCGCGCGCGCTC 2222  
 2184 GCGACGCGCGCTGCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCTC 2243  
 2223 GCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCTC 2282  
 2244 CCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCTC 2303  
 2283 GCGAAGCTGTCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2342  
 2304 GCGCGCGCTGCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCTC 2363  
 2343 CGCACCGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCTC 2402  
 2364 CCGCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCTC 2423  
 2403 GAGTTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCTC 2462  
 2424 GCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCTC 2483  
 2463 GCGTTTTCATCTCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2522  
 2484 GAGTACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCTC 2543  
 2523 AAGAACACCGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2582  
 2544 ACCAACACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCTC 2603  
 2583 GCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCTC 2642  
 2604 TCCGACGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCTC 2663  
 2643 CCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCTC 2684  
 2664 GCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCTC 2705

RESULT 8  
 LOCUS ATUCBG1 3710 bp DNA linear BCT 26-Apr-1993

DEFINITION A. tumefaciens beta-D-glucosidase (cbg-1) gene, complete, cds.  
 ACCESSION M59852  
 VERSION M59852.1 GI:142221  
 KEYWORDS beta-D-glucosidase.  
 SOURCE A. tumefaciens DNA.  
 ORGANISM Agrobacterium tumefaciens  
 Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;  
 Rhizobiaceae; Rhizobium.

REFERENCE  
 1 (bases 1 to 3710)  
 Castle, L.A., Smith, K.D. and Morris, R.O.

TITLE Cloning and sequencing of an Agrobacterium tumefaciens  
 beta-glucosidase gene involved in modifying a vir-inducing plant  
 signal molecule  
 JOURNAL J. Bacteriol. 174 (5), 1478-1486 (1992)  
 MEDLINE 92165721  
 PUBMED 1537792  
 FEATURES  
 source location/Qualifiers  
 1. 3710  
 /organism="Agrobacterium tumefaciens"  
 /db\_xref="taxon:358"  
 /cell\_line="B3/73"  
 /issue\_1b="genomic HindIII cosmid"  
 169. 2637  
 /gene="cbg1"  
 169. 173  
 /gene="cbg1"  
 181. 2637  
 /gene="cbg1"  
 /EC\_number="3.2.1.21"  
 /codon\_start=1  
 /transl\_table=1  
 /product="beta-D-glucosidase"  
 /protein\_id="AAA2082.1"  
 /db\_xref="GI:142222"  
 /translation="MIDIDLDKMTLEBOVSLSCADPTWTVATRLGVPRKITYDGN  
 GARGGSLVGVSKSPFVAIALGATWDELIERAGVALGQASKASVLAFTVNI  
 HRSGLNGRFECYSEDPALTAACAAVAVANGVOSGVATIKHFVANSERLERTSSD  
 VDERTLREIYLPPEEAVKAGVAVMSYKNGTYSSEMPMLTFLYREMFSDV  
 VMSDQGEHSHTAETINAGLIDEMPGVMSDREKTLVAAREKVAEYVARSARILL  
 LERGAEPKADIDAEHADIPEDRALLIOLAEBAVALIKNGSVPLAKASRDIOLAVIG  
 PNAASARYWGGSSARTIAHTVVSPELEGRALSNANSLRHAYGNNRLIDVSEMT  
 VEYERGRPEERPHAVETVEKEFPEEDLPSSGDLADFSARMTATVPOTGEHTIG  
 MTNAGLARLFVDGELVDYDQWTKGENEFPTARSSEORAVTLGAARYRVVEYAP  
 KASLDGINICALREGEKPLDAGIAEAVETARKSDIYLIVRGEGWDEGLDLPD  
 RLPGROELIEVAETNPVYVLOTGPIIMPMLGKRAVLQWYVGOELGNALAY  
 LFGYEPAGRLPOTPEKALTDNSATIDDPISYPODGHVRAEGLIENGVRHDTREI  
 PLFPGFELGTRFTWGPOLSGFEMGSDGLTVYDVNTIGDRAGSDVOLYVSPNA  
 RVERFEKRLRAFKALKLAPGATRAVLAIAPDLAIFYDVEAGRRAADAGKLELYAAS  
 AIDIRASVSIHLPVDHVMPE"  
 misc\_feature  
 213. 223  
 /gene="cbg1"  
 /note="active site"  
 BASE COUNT 718 a 1084 c 1168 g 740 t  
 ORIGIN  
 Query Match 11.5%; Score 343.4; DB 1; Length 3710;  
 Best Local Similarity 49.1%; Pred. No. 9,7e-36;  
 Matches 1195; Conservative 0; Mismatches 1171; Indels 69; Gaps 8;  
 355 TTGAGGCATCTTGAAGAAGCTACCCCTGCGGAGAGAGTCCATCTGCTGCTGATCG 414  
 185 TCGACGATATTCGATTAAGATGACACTGAGAGAGAGTGTGCTGCTGCGGCGCG 244  
 415 ACTTTGGGACCAAGAGGCTCCCAAGCATAGTATCCCTCTCCCTTACAGATG 474  
 245 ATTCTGGAGACGACCGTCGCGATGACGCGCTGCGCGAAGATCAAGTTACGAG 304  
 475 GCCCAACGCGCTAAGAGGACCAAGT--TCTTCAATGGCGCTCCGCGCTGCTTC 531  
 305 GCCCAATGCGCGACGCGGCGGCGGCTGCGGCGCGCTCAAGTCCGCTGCTTC 364  
 532 CTTCGCGCGAGCTGCTGCTGCTCAATCAACCAACTGCTGCGAAGAGGAGGATGA 591  
 365 CGGTGGCAATCGCGCTTGGAGCGAGTGGAGACCGGAGCTCATCGAGCGCGCGCTG 424  
 592 TGATGGGCAAGAGGACCATGCTAAGAGTGGATGATCCCTGCGCGAGTACTAACA 651  
 425 CGCTGGAGAGACCAAGGACGAGGCGCGCTGCTGCTGCGCGCGAGCGTCAACA 484  
 652 TGCAGCGCTCCCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 711  
 485 TTCACCGCTCGCGCTCAATGCGCGCAACTTGAATGTAATTCGGAAGACCGGCGCTGA 544

QY 712 CGGCGTGGAGCTCGGCTCTCATCCGCGCATTCAGACACTGAGTGCAGGCTACGA 771  
 Db 545 CGGCGCTGGCGCGCGCTATATCAATGAGCGTACAGCGAGGCTGGCGCGCACGA 604  
 QY 772 TGAAGCACTTTTGTGATGATGATGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 831  
 Db 605 TCAAGCACTTTGCTGCAACAGAGTCCGAGATGAGAGAGAGAGAGAGAGAGAGAG 664  
 QY 832 CGGAGCGGCTCTCGGTAATCTACGCACTCCGCTCCAGATTCCTGAGAGACTCC 891  
 Db 665 ATGAGCGGAGCGTGGCGCAATCTATCTGCGCCCTTTGAGAGAGCGGTGAAGAGCG 724  
 QY 892 AGCGGCTGCTTATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 951  
 Db 725 GCGTGAAGCGCGCTCATGCTCTCTACAAAGCTACAGCGCACTATACGAAATTC 784  
 QY 952 CTAAATATCTGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1011  
 Db 785 CTTGGCTGCTGAGCAAGTCTCTGCGCGAGAGATGGGCTTGGAGCGGTGCTGATG 844  
 QY 1012 ACTGCTACGCAATACAGTACAGTACAGTACAGTACAGTACAGTACAGTACAGTAC 1071  
 Db 845 ACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 904  
 QY 1072 CCGGACCTCCAGCTTCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1131  
 Db 905 CGGAGCTTGGCGGAGTGGCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 964  
 QY 1132 TATTCACAGTATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1191  
 Db 965 AGCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1024  
 QY 1132 CCTCGGAGTACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1251  
 Db 1025 CTTTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1084  
 QY 1252 CTCTCTCGGAGAGTGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1311  
 Db 1085 TCAATCGGCAACTGCTGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1144  
 QY 1312 TGGCTTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1371  
 Db 1145 TTGCAAGTCTGCTTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1204  
 QY 1372 ACCAGGCGGAGAGCTTCTGCGCACTCAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1431  
 Db 1205 TCAATGCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1264  
 QY 1432 TCAGCAAGCACTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1491  
 Db 1265 TTGCGGCGGCGCT-----GTCCAGCGCAAGAGAGAGAGAGAGAGAGAGAGAGAG 1303  
 QY 1492 CCAATTTAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1551  
 Db 1304 CGGCTGCTTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1363  
 QY 1552 TCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1611  
 Db 1364 ACTTCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1423  
 QY 1612 CGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1671  
 Db 1424 AATTTCTTGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1483  
 QY 1672 TGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1731  
 Db 1484 TGAGCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1543  
 QY 1732 GCAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1791  
 Db 1544 GCGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1603  
 QY 1792 CCGGCGATGCTCTTCTGCGGCTCGCGCAAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1851

Db 1604 AGGCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1663  
 QY 1652 AGGCAACAGCTTCAAGATTCAGATTCAGATTCAGATTCAGATTCAGATTCAGATTC 1911  
 Db 1664 CCGGAGCGGCTTACCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1708  
 QY 1912 GCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1971  
 Db 1709 CCAGCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1765  
 QY 1972 ACAGGCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2031  
 Db 1766 GCGATGCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1825  
 QY 2032 GCGGCGGCTTAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2091  
 Db 1826 TCGTGGCGGCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1885  
 QY 2092 CCGGCGGCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2151  
 Db 1886 CCGGCTGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1945  
 QY 2152 TCAATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2211  
 Db 1946 TACTGCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2005  
 QY 2212 AGGCTGCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2271  
 Db 2006 AGATGTGCTATCCCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2065  
 QY 2272 ACAACCCCTGCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2331  
 Db 2066 TCGAGCTGCGGCGGCTGCTCCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2125  
 QY 2332 TTCTCAACTTCGCGAC-----CGAGGCGGCGGAGAGAGAGAGAGAGAGAGAGAGAG 2379  
 Db 2126 TTACCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2185  
 QY 2380 TCTACGTCGAGTACAGTACAGTACAGTACAGTACAGTACAGTACAGTACAGTACAGTACAG 2439  
 Db 2186 TCTTCTGCGGCTATCCGCACTACAGATACAGATACAGATACAGATACAGATACAGATACAG 2245  
 QY 2440 ACGGCTGCTTCAACACTTTTCCCTTTCATCTCCGCTGCTCAAGAGAGAGAGAGAGAGAGAG 2494  
 Db 2246 TCGGCTTGGCTACACCGGCTTACCTGGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2305  
 QY 2495 ----CGGCAAGCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2550  
 Db 2306 GGGCGAGTGTCTTACGCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2365  
 QY 2551 AGGTGGCGGAGCTTACGCTCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2610  
 Db 2366 AGGTGGTGAAGCTCTATGATCACTCTCCCA---TCCAGAGAGAGAGAGAGAGAGAGAGAGAG 2422  
 QY 2611 AGCTCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2670  
 Db 2423 AGCTGCTGCTTGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2479  
 QY 2671 AGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2730  
 Db 2480 TGAAGATGCTCTCTCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2539  
 QY 2731 AAAAGGCTATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2765  
 Db 2540 ATGCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2824

RESULT 9  
 AB1293760 2636 bp mRNA linear PLN 09-AUG-2000  
 LOCUS  
 DEFINITION Agaricus bisporus mRNA for putative beta glucosidase (bgl gene),  
 strain D649.

ACCESSION AJ293760  
 VERSION AJ293760.1 GI:9796011  
 KEYWORDS beta glucosidase; bgl gene.  
 SOURCE Agaricus bisporus.  
 ORGANISM Agaricus bisporus  
 Eukaryota; Fungi; Basidiomycota; Hymenomycetes; Homobasidiomycetes; Agaricales; Agaricaceae; Agaricus.  
 REFERENCE 1 (bases 1 to 2636)  
 MORALES-ALMORA, P. and THURSTON, C.F.  
 Molecular analysis of the cellulolytic genes in Agaricus bisporus  
 JOURNAL Unpublished  
 REFERENCE 2 (bases 1 to 2636)  
 MORALES-ALMORA, P.  
 Direct Submission  
 Submitted (08-AUG-2000) Morales-Almora P., Microbiology, King's  
 College London, 150 Stamford Street, London, SE1 8WA, UNITED  
 KINGDOM  
 FEATURES  
 source location/Qualifiers  
 1..2636  
 /organism="Agaricus bisporus"  
 /strain="D649"  
 /db\_xref="taxon:5341"  
 /clone="PMR1615"  
 /country="United Kingdom"  
 25..2610  
 /gene="bgl"  
 /gene="bgl1"  
 25..2610  
 /function="putative hydrolysis of beta-D-glucose"  
 /product="putative beta glucosidase"  
 /protein\_id="CAC03462.1"  
 /db\_xref="GI:9796012"  
 /translation="MPSDFANMDIDKIVALTDEAIIISAGVGHHTAIRELOIP  
 AVKSDPNIGRGNHFMPTAPACLPSTAMGATFEDDLIEGLKLAELAKRSNS  
 VILATPCNIONRNPGRFSESEDPGLSGLSAYINGLOKGIKATIEHFNDRKE  
 LDRGYSIHRYTSSGDLSPVHAGRNMLHGAIVTAVRNAGIHCNPELLOKI  
 LRDCKSDVMSWRECTSYVDVNLNGLDEMPGINKMSLSVNSIOSRRTVAKK  
 VKERARKVLELVKCAQAPBILDGDELTLSDEENALMRFAESIYLLANDKKY  
 LPIIDKSLKVAIVGNNAKAIVLSGGSAALKBSEYFARPAGLKALKENPMEVYI  
 SEGARAATLPTLDFLFTESGGGRTGSMHANEHNDLSVALPTPIKTYQVDETRFI  
 SSSPAGITKWLIRMGYKLPREDCKFEFGLIAAGRAKLFVPGKLVINWTRORG  
 VERFGSESEETGVLELKAGYKLEIYDFCNVRGPAQDEETIMDSNPGIRLGAIV  
 EDPPELLNSAISKADADAVIIVGILNGDETRBNDETTLALGRDELIYOKVVAAP  
 RAIVVTOGSAITMFWDYQVSLHTYLGKATDADIAIVIRGNHNSGLSLEFPR  
 LADVPAHGFSEKGVYAEGLYVGHKQHRNIEPLRFGHLSITSNISDLRS  
 QPIVSGDFDLKATVILNTGNITGSOVVDLYIGLPTSELTHRWGLRPFKMDYK  
 PGESREVELVMDRLSVSMDEKEMVVENGAVDVRVAFTSEBVGEGDELGRFVNEKGF  
 GWRGL"  
 BASE COUNT 657 a 603 c 682 g 694 t  
 ORIGIN  
 Query Match 10.3%; Score 305.2; DB 8; Length 2636;  
 Best Local Similarity 49.9%; Pred. No. 9.8e-31;  
 Matches 1106; Conservative 0; Mismatches 1053; Indels 59; Gaps 11;

QY 569 TCTGTCGAAGAGCAGT---AGATGATGGGCAAAAGGCCATCGCTAAGAGTGC 625  
 DB 270 CCTACTGGAAGAAGTTGGCTCAGCTTCTGCTAGAGAAAGAGCTCAGATCCGCTC 329  
 QY 626 TGTATCTCGGCGGCGATATCAACATGCAAGCTCCCTCTCGGTGAGACGTGGCTTGA 685  
 DB 330 TGTATCTCGGCGGCGATATCAACATGCAAGCTCCCTCTCGGTGAGAGTTTGA 389  
 QY 686 GTGATGATGAGGATCGCTTCTGCGGGCTTGGAGCTGGGCTCTATCCGCGCAT 745  
 DB 390 AGTTTGTGGAGACCTTCTCTGCTGTATGCTTTCAGTCCATCATCAGCGAT 449  
 QY 746 TCAGACACTGAGATCGAGCTACATCAAGACATTTTGTCAATGATCAGAGACAG 805  
 DB 450 TCAGAAAGGGGATATGGGGCCACTATCAAGATTCGTGGGAAATGACAAAGAGATGA 509  
 QY 806 GCGCA-TGATGTGACAGACATGCTACAGAGAGGGCTCTCGGTAAATCTGCACTCC 864  
 DB 510 TCGAAGCTGATATGACAGATTCATTACAGACCTCTTGGGAGATTATCTCTTC 569  
 QY 865 CGTTCAAGATTGCTGTGCGAGACTCCAGCG--GGTGGCTTATGACGGCTACATG 922  
 DB 570 CGTTCACTGCTGCTCAAGAAATATGCTCTTCATGGGGGATCATACCGGCTACAA 629  
 QY 923 CATCAATGGCGTGTGTCGACGAGAACCTTAATATCTTGATGAGATGCTTGAAGA 982  
 DB 630 TGTTAAGGATTCAGATTTGCGAGATCCGAACTTACAGAAATATTTACGCGATGA 689  
 QY 983 ATGGGGTGGAGTGGCTAATCATGAGCGATGTCAGGACATACAGTACACAGAC 1042  
 DB 690 GTGAAATGAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 749  
 QY 1043 CGTTGTCAGAGGCTGAGCTGAGATGCTGAGATGCTGAGATGCTGAGATGCTGAG 1099  
 DB 750 TTTAATGCGCGGCTTACCTTGAATGCGCGCTTGAATGCGCGCTTGAATGCGCG 809  
 QY 1100 ACTCAAGTTCAAGCTTCAAGAGAAAGCCCTTATTCAGCTATGACAGAGGCTAG 1159  
 DB 810 TGTCACCGATTCATATCTAGAAAGTAAACCGCCMAAAGTAAAGAGGCTGCTG 869  
 QY 1160 GGAAGTCTTCAAGTGTGCTCAAGAGTGTGCT-----GCTCGGAGTGCAGAGAAC 1213  
 DB 870 AAAAGTCTGAGGCTTCAAAAATGCTTCAAGAGGCGCCCTGAAATTTTGGACGGCA 929  
 QY 1214 CCGGAGACAGCTGTCAACAACACCCCGAAGGCAAGGCACTCTCTCGGAGAGTTGGCA 1273  
 DB 930 TGGCTTGAAGGCACTTGTATGACAGAAAGATATGCTCATGGGAAATTTGCTGC 989  
 QY 1274 CGAGGCAATGCTGCTGTAAGAGCAAGCAAGCTTCTGCTTGAAGCAAGAG----- 1328  
 DB 990 AGACTTATGCTTCTGCTCAAGAGCAAGAAAGTCTTCTGAGCGTAAAGATT 1049  
 QY 1329 -AAGAAGCGCTGATGTCGGGCCCCCAACGACGACATACAGCGGAGGCTC 1387  
 DB 1050 AAAAAGATGCAATCGGGGTGTATATCCAAAGCGCTGTACTATCTGTGTGAGGTTTC 1109  
 QY 1388 TGCCGCACTGAGGCTTACTACGACATCACTTCTTGAAGGCTTCAAGACAGCTCGA 1447  
 DB 1110 TGTGCGCTCAAGCGCTTACTTCTGCGGCGCGGCGGTGTCAAGGCTCTGAA 1169  
 QY 1448 GAGCGCGCATCTACACCGCTGCGGCTACAC----- 1481  
 DB 1170 AGAAGTCAACCAACATCGAGGTTACTTTCAGGAAAGGCGCTAGAGCGCCAGACGTT 1229  
 QY 1482 ACCGTTCTTCCATCTAGAGGAGCAGTGTCTACGCGCGGCGGCTCCGGGATCGC 1541  
 DB 1230 ACCGACTTTGATTTGATTTGTTTCTACGAGTC--CGGCAACGAGGCTGAGAGTGGCTT 1288  
 QY 1542 TGAAGGCTTCAACGAGC-----CCGTGATACCTTACCGGACGACATGACAGCT 1597  
 DB 1289 GGCATGCTCATGAAATGACATGAGCTCTGCTTGTGCTTACCGCATTAACACAAAT 1348  
 QY 1598 CTCTTACCAAGAGCAGATGACATGAGCTGTGATCTACTACACCCCAAGGCGGACAGAC 1657



Db 1349 ATGTTCAGCAGACAAAGATTTCATCAGCATCTTCTCTCCGCTGATTAACCAAGAGT 1408  
 1658 GTGTACGCGCAGATGAGAGGACGACGACACCCGCGACAG-GACTGCACCTACAGGCTCG 1716  
 Db 1409 GAGCGCTTGTATGAAAGGTATTTTGAACCCGAGAGCGTATTTGAATTCGAAATTTG 1468  
 1717 GCGTGTGTGTGCGGACGCAAGAGGCGTACGTAGACGACAGCTGCTGTGCAACAG 1776  
 Db 1469 GATTGACCGCTCTGTGAAAGCCAGCTGTTGATGATGAAACCTGGATGATGACAACT 1528  
 1777 GCACCAAGAGGTCCCGGCGATGCTTCTTCCGCTCCGACCCGCGAGAGAGCGGCC 1836  
 Db 1529 GAGCGCTTACGTCGCGGTGATGAAATCTTGTGATGTTGATGAAAGAGAGACTGCGC 1588  
 1837 GCATCAATCTCTCAAGGCGCAACGATACAGTTCAAGATGAGATTC-----GACTCGC 1890  
 Db 1589 TTGTTGATGTAAGACCGGTGTCAGACATGATGATTTATGTGACTTCTGTATGTTGCTG 1648  
 1891 CACCACTACACCTCAAGGCGACACACATGCTCCCGCGACGCGCTCCGCTGCTG 1950  
 Db 1649 GCGCAGCTGATGATGATGAAACCGCAACATCATGATGATGATGATGATGATGATGATG 1708  
 1951 GCGGCTGCAAGGTCAATGACGACGCGCAATGCAAGTCCGCTCCGCTCCGCAAG 2010  
 Db 1709 GGGGCGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1768  
 2011 AGCAGCAGCAGGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 2070  
 Db 1769 AGCGTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1828  
 2071 ACCGCGCAGGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 2130  
 Db 1829 ACCGATACACCTTCCCTCCGCGGTACCGAGCAATGATGATGATGATGATGATGATGATGATG 1888  
 2131 CGAACCAACACCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 2190  
 Db 1889 CAATTCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1948  
 2191 AGCCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 2250  
 Db 1949 ACCAAGTTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 2008  
 2251 CGAGCTGCTTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 2310  
 Db 2009 CGAGCTATTTTGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 2068  
 2311 GCGTGCAGGACACCCGCGCTTCTCAACTTCCGACGAGGCGCGCGACGCTGATG 2370  
 Db 2069 GATTGAGAGATGTTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 2128  
 2371 GCGAGGAGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 2430  
 Db 2129 CTGAGAGCTTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 2188  
 2431 CTTTGGCGAGGCTGCTCTCAACACTTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 2488  
 Db 2189 CGTTGGCGAGCTTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 2246

# RESULT 10 REU80928\_2/c

Sequence split into 4 fragments LOCUS REU80928 Accession U80928  
 Fragment Name Begin End

REU80928\_0 1 110000  
 REU80928\_1 100001 210000  
 REU80928\_2 200001 310000  
 REU80928\_3 300001 371255  
 Continuation (3 of 4) of REU80928 from base 200001 (U80928 Rhizobium etli strain CFN42)

Query Match 9.1%; Score 270; DB 1; Length 110000;  
 Best Local Similarity 48.0%; Pred. No. 2.7e-26;

Matches 1116; Conservative 0; Mismatches 1135; Indels 75; Gaps 9;  
 348 ATTGATGTTGAGGCATCTTGAAAGCACTACCTTGCCGAGAGGATGATGCTGCTGCT 407  
 Db 81149 ATTGATGAAAGCATTGCTGATTAATATGACTTGGGAGAGGCTGCTGCTGCTGCT 81090  
 408 GGTATGCACTTGTGACACAAAGGCTCTCCCAAGATGATGATGCTCTCTCTCTCTCTCT 467  
 Db 81089 GGCATATGCTTCTGCTTTTACCGCCCATGACCGCTTAGGATAGGAGATATGCTGCTG 81030  
 468 ACATATGCCCCCAAGCGCTGAAAGGACCAAGTTCT--TCAATGCGCTCTCTCTCTCTCT 524  
 Db 81029 ACCGACGCCCCCAAGCGCGCGCGCGCGCGCGCGCTTCTTCTGAGATGATGACGCG 80970  
 525 TCGTCTCTCTGCGGACGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 584  
 Db 80969 GCTTTCCTGCTGCGGATGCGCATGCGCATGCGCATGCGCATGCGCATGCGCATGCGCAT 80910  
 585 GGTATGATGATGAGGCAAGAGGCGCATGCTTGAAGATGCGCATGCTGCTGCGCGCTGCT 644  
 Db 80909 GCGATGCTCCTAGGTGACGAGGTTCTTCAAAAGTGCCACAGTCTGTTAGCCCGCCAC 80850  
 645 ATCAATGATGAGGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 704  
 Db 80849 GTTACATCCAGGCTGATGCTGACCAAAAGGCGCGCACTTGTGATGCTTCTGAGATGCG 80790  
 705 TTCTTGGCGGCTTGGAGCTGCGGCTCTCATCCGCGCATTTAGAGCACTGAGATGAG 764  
 Db 80789 ATCTGACGCGGAGACTGCTGAGTGGCTATATGAAAGTCTGAGTCCAGAAATGCTGCT 80730  
 765 GCTATGATGAGCACTTTTGTCAATGATGATGAGAGGACAGCGCATGATGATGATGATG 824  
 Db 80729 GCGACGATTAATGATGCTGCGCAAGGATGCGAATGATGATGATGATGATGATGATG 80670  
 825 ATGCTACGAGGCGGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 884  
 Db 80669 GATATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 80610  
 885 GACTCCAGCGGCTGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 944  
 Db 80609 CGGCAAAAGGCTGCGCGCGGATGATGATGATGATGATGATGATGATGATGATGATGATG 80550  
 945 GAAACCTCTAATATCTTGTATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1004  
 Db 80549 GAAAGCATTGCTGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 80490  
 1005 ATGAGGAGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1064  
 Db 80489 ATGTCGACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 80430  
 1065 GAGATGCGCGGACCTTCACGCTTCCGAGGAGAAACACTCAAGTTCAAGCTTCAACGGA 1124  
 Db 80429 GAAATTCGCGGCGCGAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 80370  
 1125 AAGCCTTATCCAGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1184  
 Db 80369 GAGGTTAGTGTGAGACGATCCGCGCTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 80310  
 1185 TGTGCTGCTCTCGAGTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1244  
 Db 80309 ACCGCGCTATCAAGCATATGCGCGAGTTTAAAGAGTACCGCATGATCAACCGAATCAT 80250  
 1245 ACAGGACCTCTCTCTCGAGAGTTGGCAACGAGGCGATGCTGCTGTAAGAGAGAGAGAG 1304  
 Db 80249 AGGCGATTGATCAAGCGCGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGATC 80190  
 1305 AAGCTTCTGCTGCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1364  
 Db 80189 -----CTGCACTTGTGCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 80136  
 1365 GCCACATACAGAGGCGGAGGCTCTGCGGCACTGAGGCGCTTACGAGAGTACGCTCCCTT 1424  
 Db 80135 GCGCAGGATGATGAGAGGCGGCTGCGGCACTTGAACCTTCACTATGATCATGATGCTG 80076





KEYWORDS beta-glucosidase; elongation factor G; fusa; succ; succinyl-coa synthetase alpha chain; succinyl-coa synthetase beta chain; succ; TPP-requiring enzyme; transfer-RNA-Leu; TTA Leu codon.

SOURCE Streptomyces coelicolor A3(2).

ORGANISM Bacteria; Firmicutes; Actinobacteria; Actinobacteridae; Actinomycetales; Streptomycinae; Streptomycetaceae; Streptomyces.

REFERENCE 1 (bases 1 to 37445)  
Redenbach, M., Kieser, H.M., Denaplatte, D., Eichner, A., Cullum, J., Kinash, H. and Hopwood, D.A.  
A set of ordered cosmids and a detailed genetic and physical map for the 8 Mb Streptomyces coelicolor A3(2) chromosome  
Mol. Microbiol. 21 (1), 77-96 (1996)

TITLE

JOURNAL

MEDLINE 97000351  
8843436

PUBMED 2 (bases 1 to 37445)  
Seeger, K.J. and Harris, D.  
Unpublished

REFERENCE 3 (bases 1 to 37445)  
Parkhill, J., Barrell, B.G. and Rands, M.A.  
Direct Submission  
Submitted (01-JUL-1998) Streptomyces coelicolor sequencing project, Sanger Centre, Wellcome Trust Genome Campus, Hinxton, Cambridge CB10 1SA E-mail: barrell@sanger.ac.uk Cosmids supplied by Prof. David A. Hopwood. [3] John Innes Centre, Norwich Research Park, Colney, Norwich, Norfolk NR4 7UH, UK  
On May 9, 2002 this sequence version replaced gi:3288600.

AUTHORS

REFERENCE

TITLE

JOURNAL

## COMMENT

## Notes:

Streptomyces coelicolor sequencing at The Sanger Centre is funded by the BBSRC.

Details of S. coelicolor sequencing at the Sanger Centre are available on the World Wide Web.

(URL: [http://www.sanger.ac.uk/Projects/S\\_coelicolor/](http://www.sanger.ac.uk/Projects/S_coelicolor/)) CDS are numbered using the following system eg SC7B7.01c. SC (S. coelicolor), 7B7 (cosmid name), .01 (first CDS), c (complementary strand).

The more significant matches with motifs in the PROSITE database are also included but some of these may be fortuitous. The length in codons is given for each CDS.

Usually the highest scoring match found by fasta -o is given for CDS which show significant similarity to other CDS in the database.

The position of possible ribosome binding site sequences are given where these have been used to deduce the initiation codon. Gene prediction is based on positional base preference in codons using a specially developed Hidden Markov Model (Krogh et al., Nucleic Acids Research, 22(22):4768-4778(1994)) and the Frameplot program of Bibb et al., Gene 30:157-66(1984) as implemented at <http://www.nih.gov/jp/jun/cgi-bin/frameplot.pl>. CAUTION: We may not have predicted the correct initiation codon. Where possible we choose an initiation codon (atg, gtg, ttg or att) which is preceded by an upstream ribosome binding site sequence (optimally 5-13bp before the initiation codon). If this cannot be identified we choose the most upstream initiation codon.

IMPORTANT: This sequence MAY NOT be the entire insert of the sequenced clone. It may be shorter because we only sequence overlapping sections once, or longer, because we arrange for a small overlap between neighbouring submissions. Cosmid 8A6 lies between 3f9 and 1f2 on the AseI-A genomic restriction fragment.

## FEATURES

## source

## misc\_feature

## stem\_loop

## gene

## CDS

```
1. 37445
/organism="Streptomyces coelicolor A3(2)"
/strain="A3(2)"
/db_xref="taxon:100226"
/clone="cosmid 8A6"
2. 107
/note="nominal overlap with st8A6 from 1 to 106
nominal overlap with cosmid 3f9 from 19725 to 19830"
657. 711
/note="hairpin loop with 23 bp stem"
complement(730. 2160)
/gene="SC06581"
complement(730. 2160)
```

## gene

## CDS

```
/gene="SC06581"
/note="SC8A6.02c, probable transmembrane transport
protein, len: 476 aa; some similarity to e.g. TR_051330
(EMBL:04007 5) oxalate:formate antiport protein from
Oxalobacter formigenes (418 aa), fasta scores: opt: 226
z-score: 230.1 E(): 1.4e-05, 24.1% identity in 436 aa
overlap"
/codon_start=1
/transl_table=1
/product="putative transmembrane transport protein"
/protein_id="CAA19774.1"
/db_xref="GI:3288602"
/db_xref="SPTREMBL:087836"
/translation="MGQALARSFDRAGAIADTTTDTVTRVYATREYTRNGRYRVGE
SDIDLGRKRKRWVILPWIGMGITSSAEAFASAEEDTLEAQMSSGSIYMMATAVF
COAAVAFPARLRESGKLPAARMAMLSAGTGLGLSLAFAPVAFVAFGFGVSGNG
AGWYATCVNMVGRKWPERRGKRTGFGVNGGFAVSGVPFIFGKYDTSNFRVLSA
GVFLAMVATAGFYFRDPKRMMPAAVDLPNPPADRRARSLKKNPAAVKOYSPMEAW
KTRGVALLMEFCLACTSGVNIPTGIAFOVDIGEEGFAVYAAAMSKATVNGRGVI
GMSIDLYGRKOCILYCATILGLAOPFTIISAEKKNIPLVYSATISFGCGAIFPWEA
ALPADPFGENNNATNTGMYTSKLVSGLGAGMGSVVGVWYNGAFSLAGSISIFAGF
VALSLRPGKPKKRVTPNPPGEG"
complement(2233. 4377)
/gene="SC06582"
complement(2233. 4377)
/note="SC8A6.03c, unknown, len: 714 aa; similar to many
hypothetical proteins including several from each of the
archaeons Archaeoglobus fulgidus TR_028341 (EMBL:AE000970)
conserved hypothetical protein AFI1938(673 aa), fasta
scores: opt: 960 z-score: 1289.8 E(): 0.35, 64% identity in
693 aa overall and Pyrococcus horikoshii D1028624
(EMBL:AB009518) hypothetical protein P1C016 (460 aa),
fasta scores: opt: 1280 z-score: 1246.9 E(): 0.44, 3%
identity in 461 aa overlap. Contains PS00216 Sugar
coelicolor hypothetical protein SC9B10.09 (E(): 6.1e-15,
33.9% identity in 375 aa overlap)"
/codon_start=1
/transl_table=1
/product="conserved hypothetical protein"
/protein_id="CAA19775.1"
/db_xref="GI:3288603"
/db_xref="SPTREMBL:087837"
/translation="MAEDRLAVRGLIADVRGEGRTALTAPRGYLAARYGAVGEE
IARDVDEAFACARCGPVMYKIVTPDILHRTDAGCVVGEADVRAARFRTDNA
RAYDSARIREGVOIBLPOGEVITGAVTDPTGKVAAGLGLVLEALCORTERLA
PVADDEALSLMDSIRAEVLRGVRQAGDRAVAEQIRVSELVADPEIAEDLPN
VIATPEGAAVADIRVILATDPEKERRRTREILSMRKLMPASVAVIAGSGEGLI
GNSVRNLVDGGEAGEIHPVNRKADIDIGRKAYKSVTVGPDVAFAIPAREVAAA
LEEVGRKRIPNPAVLIPSGFAETGEALODEIVAVERGVRLLGNITGYSTMODIC
AFPCPVYDQKGVALLTSGGCGIIMLTPARTKGVSAIYGCKSLDDEDLITWF
GDPHTECIAMLELDKGRARVEAARTRPKPVYKAGRTAAGAAASHTGALA
GDDAYVDILRAGVTRAPGLEMELVRALEVLPAPEGDNVVIITGAGSGVLSLA
VTDGLSLMETIPDPDRAFRATIPFGAAGNPDVITGEPSTYEATRTILGIEPRH
ALVITGVHTIVPDPVFAELARVAFEEERREKVEPVVAAISLAGDVEVEACQYLFEG
VVAAYTEKTEPAVVGARYMARAGLIGGGS"
complement(3340. 3387)
/gene="SC06582"
/note="PS00216 Sugar transport proteins signature 1"
complement(4382. 5614)
/gene="SC06583"
complement(4382. 5614)
/gene="SC06583"
/note="SC8A6.04c, probable transferase, len: 410 aa;
highly similar to TR_006644 (EMBL:U82167) formyl-CoA
transferase from Oxalobacter formigenes (428 aa), fasta
scores: opt: 95 z-score: 1732.0 E(): 0.50, 7% identity
in 428 aa overlap. Also similar to many dehydratases e.g.
CAIB_ECOLI L-carbit line dehydratase (EC 4.2.1.89) (405
aa), fasta scores: opt: 285 z-score: 317.7 E(): 1.9e-10,
26.2% identity in 427 aa overlap"
/codon_start=1
```

## misc\_feature

## gene

## CDS



QY	1709	CGACACTGGCGCTGTGTGTGCGGCGACCGGCAAGGCGTACGTACGACACGACCGTCTG	1768
Dp	28985	CTTGAGC-CCCGCCTGGTGTGCGGGCGCGACACCGCTCGAGATCCGCGCCGGGTGTCTC	29043
QY	1769	CGACAAACGCCACCAACAGAGTGTCCCGGCGATCCCTTTCTTCGGCTCCGCGACCCGCGAGGA	1828
Dp	29044	CCGAGACGGGGGCTCCGTGGTCCCTGGGCGGTGGCCGGTTCGGCGGAGTGAAGCTTGACGAC	29103
QY	1829	GACGGGCGCGATCAATCTGCTCAAGGGCAACAG-----	1862
Dp	29104	GGAAGGACGACACCTGCTGGAGGGGAGACTCCCGCGTCCACGACGATCCGGCGGTGAT	29163
QY	1863	-----TACAGTTCAAGATCGAGTTTGGGCTCCGACACCACTACACCTT	1906
Dp	29164	GCACGTCAACCCCGCCGCCAATGACGCACACCGCGCACTCAACCGCGGGGCAACACCT	29223
QY	1907	CNAGGGCGACA-----CCATGCTCCCGGGCCAGGGCTTCCTCCGCGTGGCGG	1954
Dp	29224	GCTGTGTGGCCCGGCGGAGGTGACACCCGGCACCGGCGGGCGACCGTCTCTCGGGG	29283
QY	1955	CTGCAAGGTCAATTGACGACACCGGCCAATTCGAAAGTCGCTGCGCCCTCGGCAAGAGCA	2014
Dp	29284	CCCGCCGCCCCGAGCTGACCGCGTGTCCGCGCAAGGCTCCGCGCGGCGGTGTGCGG	29344
QY	2015	CGACACGGTCAATCTGCGCGGGGCTTTAACGCGCATGGGAGACGAGAGGCGCGACCG	2074
Dp	29344	GGAAGCGCGGGTGTGTGCTGTGCGCACCGACGACGAGGAGATGCGAGGGGCTACGACG	29403
QY	2075	CGCGAGCATGAAGTCTCCCGCGCTGCTGGACAGAGTCATTTCCGACGTTGGCGCGCGGA	2134
Dp	29404	TACGACACTGGCGTGTGGCGCCACCCACGAGACCGCTGGTTCGCGCGGTGGGGCGGCA	29463
QY	2135	CCCAAAACACGTCGTCGTCATTCGAGACGGGCGACCCCGAGAGATATGCTCTGGCTCGAGC	2194
Dp	29464	CCCGGACACGTCGTCGTCGTCATTAAGAGGGCGGGCCCGGTGGAGATGCGTGGGGAGCA	29522
QY	2195	CACGCCGCGTCAATCCAGGCTTGTAACGGCGGCAACGAGACGGGCAATCTCAATTGGCA	2254
Dp	29524	GCGCGGTGGGTGTGCTGTGCTGTCGTTGTTCCGGACAGAGGGGCGGGTGGACTGGCGCA	29583
QY	2255	CGTGTCTTTGGCAGTACACAAACCCCTCGGGGAAAGTGTTCCTTCAGACTTCCCAAGCCCT	2314
Dp	29584	CGTGTCTTGGGAGACCGCCGAGCGAGCGGGGGAGAGACTGCCCAACACTGTGGC---GGCGT	29644
QY	2315	CGAGGACAAACCCCGCTTTCTCAATCTCCGCAACGAGCGGGGGCACGCTGTACGGCA	2374
Dp	29641	CCTGCGCGACCCCGCGGTACCCCGCACCGCCCGCCACGCGCGCGCTCTCGACTCGAGCA	29700
QY	2375	GGAGGTGTACGTCGGGATACAGGTACTACGATGTTGGCGCAAGGACGTCAATTTCCACTT	2434
Dp	29701	GGGACTGACCTCGGTACCGGGGGCTGGCTGGCGCATCAACGCAAGCCGCGCTACTGTGTT	29766
QY	2435	TGGCGAGCGCTGTCTACACCACTT-----TTGCTTTTCCAAATCTCTCGTGTCTC	2485
Dp	29761	CGGAGACGGGTCTCGGTACACAGACGAGGGGTACGAGAGAGCTGACCGCTCCGCGGTGAC	29820
QY	2486	TCACAAAGACCGCAAGCTGAGCGTGTCTCTCCGTGAAGAACACCGGCTCCGTCGGCG	2545
Dp	29821	CCGGCGGGGACACGGCTCTACCGCTGTGGCGGTGGGGTGGCGCAACCGGTGTGCGAGGGG	29886
QY	2546	CGCACAGGTGGCCAGGTCTACGTCAAGGCCCTCCAAAGGAGGCGCAAGATTAACCGCCCTG	2605
Dp	29938	GCCTTGGCTCCCGGGTACACGCGCGTGTGGCGCGCGCCCGGGGGAGAGCGGTACGCGGAC	29997
QY	2666	AATGAGGAGCAGGAGAAATAGCTCCCTCGATTAATTTGATGAGAGGC	2712
Dp	29998	GCTGCGCTCTCCGCGCGGGCCCTTGGCGCACTGTGTGCTGGCGGGAGC	30044

LOCUS	DEFINITION	ACCESSION	VERSION	KEYWORDS	SOURCE	ORGANISM	REFERENCE	AUTHORS	TITLE	JOURNAL	FEATURES	BASE COUNT	ORIGIN	Query Match	Best Local Similarity	Matches 1028;	Conservative	Score 184.4;	DB 6;	Length 3241;	Pred. No. 5.8e-15;	Mismatches 956;	Indels 150;	Gaps 12;
AR173232	Sequence	AR173232	AR173232.1	GI:17912723	Unknown.	Unclassified.	1 (bases 1 to 3241)	Belach, M.C. and McDaniel, R.	Streptomyces nambiensis	Patent: US 6303767-A 11-16-00; 2001;	Location/Qualifiers	520 a	1174 c	1125 g	422 t									
AR173232	Sequence	AR173232	AR173232.1	GI:17912723	Unknown.	Unclassified.	1 (bases 1 to 3241)	Belach, M.C. and McDaniel, R.	Streptomyces nambiensis	Patent: US 6303767-A 11-16-00; 2001;	Location/Qualifiers	520 a	1174 c	1125 g	422 t									
AR173232	Sequence	AR173232	AR173232.1	GI:17912723	Unknown.	Unclassified.	1 (bases 1 to 3241)	Belach, M.C. and McDaniel, R.	Streptomyces nambiensis	Patent: US 6303767-A 11-16-00; 2001;	Location/Qualifiers	520 a	1174 c	1125 g	422 t									
AR173232	Sequence	AR173232	AR173232.1	GI:17912723	Unknown.	Unclassified.	1 (bases 1 to 3241)	Belach, M.C. and McDaniel, R.	Streptomyces nambiensis	Patent: US 6303767-A 11-16-00; 2001;	Location/Qualifiers	520 a	1174 c	1125 g	422 t									
AR173232	Sequence	AR173232	AR173232.1	GI:17912723	Unknown.	Unclassified.	1 (bases 1 to 3241)	Belach, M.C. and McDaniel, R.	Streptomyces nambiensis	Patent: US 6303767-A 11-16-00; 2001;	Location/Qualifiers	520 a	1174 c	1125 g	422 t									
AR173232	Sequence	AR173232	AR173232.1	GI:17912723	Unknown.	Unclassified.	1 (bases 1 to 3241)	Belach, M.C. and McDaniel, R.	Streptomyces nambiensis	Patent: US 6303767-A 11-16-00; 2001;	Location/Qualifiers	520 a	1174 c	1125 g	422 t									
AR173232	Sequence	AR173232	AR173232.1	GI:17912723	Unknown.	Unclassified.	1 (bases 1 to 3241)	Belach, M.C. and McDaniel, R.	Streptomyces nambiensis	Patent: US 6303767-A 11-16-00; 2001;	Location/Qualifiers	520 a	1174 c	1125 g	422 t									
AR173232	Sequence	AR173232	AR173232.1	GI:17912723	Unknown.	Unclassified.	1 (bases 1 to 3241)	Belach, M.C. and McDaniel, R.	Streptomyces nambiensis	Patent: US 6303767-A 11-16-00; 2001;	Location/Qualifiers	520 a	1174 c	1125 g	422 t									
AR173232	Sequence	AR173232	AR173232.1	GI:17912723	Unknown.	Unclassified.	1 (bases 1 to 3241)	Belach, M.C. and McDaniel, R.	Streptomyces nambiensis	Patent: US 6303767-A 11-16-00; 2001;	Location/Qualifiers	520 a	1174 c	1125 g	422 t									
AR173232	Sequence	AR173232	AR173232.1	GI:17912723	Unknown.	Unclassified.	1 (bases 1 to 3241)	Belach, M.C. and McDaniel, R.	Streptomyces nambiensis	Patent: US 6303767-A 11-16-00; 2001;	Location/Qualifiers	520 a	1174 c	1125 g	422 t									
AR173232	Sequence	AR173232	AR173232.1	GI:17912723	Unknown.	Unclassified.	1 (bases 1 to 3241)	Belach, M.C. and McDaniel, R.	Streptomyces nambiensis	Patent: US 6303767-A 11-16-00; 2001;	Location/Qualifiers	520 a	1174 c	1125 g	422 t									
AR173232	Sequence	AR173232	AR173232.1	GI:17912723	Unknown.	Unclassified.	1 (bases 1 to 3241)	Belach, M.C. and McDaniel, R.	Streptomyces nambiensis	Patent: US 6303767-A 11-16-00; 2001;	Location/Qualifiers	520 a	1174 c	1125 g	422 t									
AR173232	Sequence	AR173232	AR173232.1	GI:17912723	Unknown.	Unclassified.	1 (bases 1 to 3241)	Belach, M.C. and McDaniel, R.	Streptomyces nambiensis	Patent: US 6303767-A 11-16-00; 2001;	Location/Qualifiers	520 a	1174 c	1125 g	422 t									
AR173232	Sequence	AR173232	AR173232.1	GI:17912723	Unknown.	Unclassified.	1 (bases 1 to 3241)	Belach, M.C. and McDaniel, R.	Streptomyces nambiensis	Patent: US 6303767-A 11-16-00; 2001;	Location/Qualifiers	520 a	1174 c	1125 g	422 t									
AR173232	Sequence	AR173232	AR173232.1	GI:17912723	Unknown.	Unclassified.	1 (bases 1 to 3241)	Belach, M.C. and McDaniel, R.	Streptomyces nambiensis	Patent: US 6303767-A 11-16-00; 2001;	Location/Qualifiers	520 a	1174 c	1125 g	422 t									
AR173232	Sequence	AR173232	AR173232.1	GI:17912723	Unknown.	Unclassified.	1 (bases 1 to 3241)	Belach, M.C. and McDaniel, R.	Streptomyces nambiensis	Patent: US 6303767-A 11-16-00; 2001;	Location/Qualifiers	520 a	1174 c	1125 g	422 t									
AR173232	Sequence	AR1																						

QY 1264 AGTTGGCAACGAGGACATCGTGTCTGAAGAACGAGAACATGTTGCTTGAGCA 1323  
 DB 1841 CGACTCCGGGGCCCGCCCGCCGCTGACAGAGCGGGGCCAGCGGTGTCCGCAAGG 1900  
 QY 1324 AGAAGAGAGAGAGCTGATTTGCGGCCCAACGCCAGCAGGCCACATACAGCGGGAG 1383  
 DB 1901 TCGCCGAGAACGCGCGGTGTCTCTGCGCAAGAGGGGCCCGCTGCGCGGTG 1960  
 QY 1384 GCTTCCGCACTAGGCGCTTCTACGAGTACTCTCTTGAGCGCTTACAGCAAGC 1443  
 DB 1961 ACGCCGCAAGAGATCGCGCTGATCGCGCCGAGCGCGGTGAC-CCCAAGGTACCGGC 2019  
 QY 1444 TCGAGACGCGCGCATCTACACCGCTGCGGCTTACACCGCTTCTCTTCAAGCG 1503  
 DB 2020 CTGGGCAAGCGCGCTGCTGCGGCTGCGGCGCGCGCGCTGACACATCAAGCGC 2079  
 QY 1504 AGCACTGCTTACGCGCCGACGCGCTCGGGGCAATGCGCTGAGAGGCTTCAAGCGCC 1563  
 DB 2080 CGCGCGGCGCGGTGCTGCGAGTACGAGACGCGGTGAGAGACCTTGGGAGCGCG 2139  
 QY 1564 CTGGTACCGCTTACCGCGAGCATTTGACGAGCTTTTACCAAGAGGAGCATGCAAC 1623  
 DB 2140 ATCCGCGGCGCGAGCTAGC-----CGGCGTTCAACG 2173  
 QY 1624 TGGTGGACTACTACCCCAAGCGCGAGACAGTGGTACCGCGCATGAGAGGCAAGT 1683  
 DB 2174 AGGGCGACACAGCTGAGCGCGGAGAGCGGGGCGCTGTACGAGCGACGCTGACCGTGC 2233  
 QY 1684 ACACCGCGGAGAGAGATCTGACCTTACGAGCTCGGCTGTGCTGCGCGCACGCGCAAGG 1743  
 DB 2234 CGCGCGAGGCGAGTACGCGCATCTCGTCAAGGCGCGCGGTGACCGCGGTGCG- 2292  
 QY 1744 CGTACGTAGACGACAGCTGCTGCTGACACAGCGCACAGGATCCCGCGCATGCT 1803  
 DB 2293 -----CTCGGCGACCCACA 2305  
 QY 1804 TCTTGGCTCGCGCACCGCGAGAGAGAGCGCGCATCATCTGCTCAAGGGACACAGT 1863  
 DB 2306 CCATGAGAGCGCGTACGCTGACGCGAGGAGAGCGCGCTCTCAAGTACGACAGG 2365  
 QY 1864 ACAATTTCAATGATGAGTTCGCTCGCGCACCGACCTTACACCTTCAAGGGCGACACATCG 1923  
 DB 2366 GACGCGACAGCT-----CAGATCTCGGCTTCGCGATGA 2401  
 QY 1924 TCCCGCGGACGCGCTCCCTCGCGCTGCGGCTGCAAGTATGACAGCAGCGCGAA 1983  
 DB 2402 GCGCGACCGCGCTCTCTGAGAGCTGGGCTGGTGACGCGCGAGGACCGCGCGACGA 2461  
 QY 1984 TCGAAAAGTCCGTGCGCTCGCCAGAGAGACAGCAGGTCATCATCTGCGGGGCTTGA 2043  
 DB 2462 TCGCGAAGCGCGTGAAGTGGCGGAGAGCGCGCTACGCGCATGCTGTTGCG- 2514  
 QY 2044 ACGCGGATGAGAGAGCGAGGCGCGCGCGCGGAGCATTAACCTCCCGCGCTGCTGG 2103  
 DB 2515 --TACGAGAGAGCGACGAGGAGCGCTGCGACCTGCGTCTGCGGGTACGCGAGG 2572  
 QY 2104 ACAAGCTATTGCGGACGCTGCGCGCGCGGAGAACCAACACCGCTGCTCATGACAGG 2163  
 DB 2573 ACAAGCTATTGCGGCGGCTGCGCGCGAGAGAACCGAGACAGATGCTGCTCTCAACCG 2632  
 QY 2164 GCACCGCGGAGAGATGCGCTGCGCTGCGACGCGCGCGCGCTGATCAGGCGCTGTAAG 2223  
 DB 2633 GTTCTGCTGCTGATGCGCTGCGCTGCGACAGCAGCGCGCGCTGCGACATGCTGATCC 2692  
 QY 2224 GCGGAGAGAGAGCGGACATCTGCGCGAGCTGCTTGGGAGCTTACACCGCTCGG 2283  
 DB 2693 CGGCGAGAGCGGCGCGCGAGCGCGCGCTGCTTACGAGTGAAGCGGAGCG 2752  
 QY 2284 GCAAGCTTCCCTAGCTTCC-----CGAAGCGCTGCGAGAGAACCGCGCTTCTCA 2337  
 DB 2753 GCAAGCTACGAGAGCTTCCCGCGCGCGAGAACACGACAGCGCGCTGCGCGGACCGGA 2812

QY 2338 ACTTCCGACCGAGGCGCGCGCGACGCT---GTACGCGAGAGAGCTTACTGTCGGGTACA 2394  
 DB 2813 ACGCTACCGCGCGCGCTGACACACAGAGAGACGAGAGGCGCATTCACAGCTGGGTACC 2872  
 QY 2395 GGTACTGAGATTGCGCGACAGAGAGCTCAATTTCCCTTTGGCCACAGCGCTTCTTACA 2454  
 DB 2873 GCTGCTTGCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2932  
 QY 2455 CCACTTTGCTTTTCCAACTCTCCGCTGTCACAGAGAC---GGCAAGCTGAGGCTGT 2511  
 DB 2933 CTTGCTTACGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2992  
 QY 2512 CCTCTCTCGTGAAGAACACCGCGCTCGTGGCGCGCACAGAGTGGCCAGCTTCTACGTCA 2571  
 DB 2993 CGGTACAGGTGGCGACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3052  
 QY 2572 AGCCCTTCCAGAGCGCGCGAGATTACCGCGCTTCAAGAGCTTCAAGGCTTTCGAAAG 2631  
 DB 3053 GCGGAGACCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3112  
 QY 2632 TCGAAGTGCAGCGCGCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2665  
 DB 3113 TCGCGCTCGCGCGCGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3146  
 RESULT 14  
 AF521878  
 LOCUS 17665 bp DNA linear BCT 29-JUL-2002  
 DEFINITION Streptomyces narbonensis desosamine biosynthetic gene cluster,  
 partial sequence.  
 ACCESSION AF521878  
 VERSION AF521878.1 GI:22001230  
 KEYWORDS  
 SOURCE Streptomyces narbonensis.  
 ORGANISM Streptomyces narbonensis.  
 Bacteria; Actinobacteria; Actinobacteriales; Streptomyces.  
 Streptomyces; Actinobacteria; Streptomyces.  
 REFERENCES  
 1 (bases 1 to 17665)  
 Bate,N. and Cundliffe,E.  
 The desosamine biosynthetic cluster of Streptomyces narbonensis,  
 producer of narboxycin  
 JOURNAL Unpublished  
 REFERENCE 2 (bases 1 to 17665)  
 Bate,N. and Cundliffe,E.  
 Direct Submission  
 JOURNAL Submitted (14-JUN-2002) Biochemistry, Leicester University,  
 University Road, Leicester LE1 7RH, UK  
 FEATURES  
 source  
 1..17665  
 /organism="Streptomyces narbonensis"  
 /db\_xref="taxon:67333"  
 <1..528  
 /gene="nbma"  
 <1..528  
 /gene="nbma"  
 /note="terminal thioesterase I (TEI) domain of type I  
 polyketide synthase"  
 /codon\_start=1  
 /transl\_table=11  
 /product="Nbma"  
 /protein\_id="AAM88350.1"  
 /db\_xref="GI:22001231"  
 /translation="AVLRPAAGDAPVLLGSHGALLAHELAPRERAHGTAPAGIVL  
 DYPGHOEPLEVMROLEGFLFAELPEPMSDARLMAGRYARFLADRPERSAPVL  
 IYRASERIGDMEERGDWRHMDLPHSVADVDPGHFTMMRDHAPAVAVAVLAMDGIE  
 RPTAAGRTAGADQ"  
 525..1370  
 /gene="nbmb"  
 525..1370  
 /gene="nbmb"  
 /note="thioesterase II (TEII)"  
 /codon\_start=1  
 /transl\_table=11



/product-"Nbmb"  
/protein\_id="AA08351.1"  
/db\_xref="GI:22001232"  
/translation="MTDRVLNVDSSIMLRFRHSPNSAVRLCLPFIAGGASAYFRFS  
EELHPSVEALSYQYGRQDRRAEPCLESEVLEAEHVAAETAMNREGRLAEFGHSLGA  
SVAYEARILTEQHRGVRPEGLVYSGRSLAPDLRVLQDLDRFLAEIROLGSDGER  
FIQDDLMRLTY.PALRSQYKAETYEHRPSATLACPVNALAGDRPKAPLNEVAMR  
HTGSPFLCRAYSGGHFTLNDQNHGICNDISDLILTYRCAPDSRVQOPTRLLEGAAKR  
WQMPR"  
/gene="nbmb"  
/gene="nbmb"  
1367..2713  
/note="putative NDP-deoxyhexose 3,4-isomerase"  
/codon\_start=1  
/transl\_table=11  
/product-"Nbmc"  
/protein\_id="AA08352.1"  
/db\_xref="GI:22001233"  
/translation="MTDDPARALTQPLRTYSAVADPELAGHILLETGRIHMHANG  
DPTAALRGADDPYATERYVRARGLTSPGSGWTADHAAASVLCSTERGCGAD  
GVPPQOVLSTGEGCPLEHQAARAAVPGAGGRRAVEVHRIIEGLAPDPSA  
SYAFELGVGVYAVGAAAVLGPAGRGFEALVERLRPLSDSLAPQSLRTYRA  
ADGALALFLLAEAGSATDSDSPSAPSSSGPTAPSSGPTAPSSGV  
AGRPDGLSALGVTAAVOLTGNVALLAHPEOMRELCAFCIAAAVEETLRYP  
VOLDARARVETELAGRRLPAGAHVVLTAGTGRDGFETDPERFLPADTAHLAL  
HPAGTGPVASLRLDAVALRTLAGRFGLRAGVYLRPRAPVGRGLSVVGS"  
2782..4062  
/gene="nbmd"  
/gene="nbmd"  
2782..4062  
/note="desosaminyl transferase"  
/codon\_start=1  
/transl\_table=11  
/product-"Nbmd"  
/protein\_id="AA08353.1"  
/db\_xref="GI:22001234"  
/translation="MRYLTSPAHHTHYGLVPLAWALLAAGHEVVASQPLTDTIT  
GSGLAAPVGTDLIHEYRYRMAGEPPNHPATAPDARSEPLDMOHAGIDILAPY  
PYLLANDSMIDILVCFARSMQPLVIMETPTAGVAAVGTGAHARVIMGPDWS  
ARRFVALRROPREHREDPTAEMLTWTLDRTGASFEELITQOFTYDPPSLRDT  
GLPTGVQVYVNGTSVPPMLAEPPAPRVCLTLGVASREVLSGDSQADLLEALA  
LDLIELVATLDAGRAEIRSYPKHRTFDVPMHALPSCSAIVHGGAGTYATAVIN  
AVPOVMAELMDAPVAKARAAVQAGFFLPAGILTPQAVDAVVRILLDPSVLAAGR  
LRRETHRPAGIVPELERLAQHRRPADAR"  
4151..4864  
/gene="nbmb"  
/gene="nbmb"  
4151..4864  
/note="N,N-methyltransferase"  
/codon\_start=1  
/transl\_table=11  
/product-"Nbme"  
/protein\_id="AA08354.1"  
/db\_xref="GI:22001235"  
/translation="MYEVDHADVYDLFLGRKDYAAEASDIADLVTRTPPEASLLD  
VAGGTGHEHTEKEFGPTAGLESEMILTHARKRLPDATLHOGMDRDFLGRFSAY  
VSMFSVGYLTETLEDAVAFNAHLEPGVYVVEPMWPEPTADWGSADVRRDG  
RTVAVRSHSVYDGDATREYHFTYADRCGRVHSDVHLITLFRAYEAAFAAAGR  
VEYLEGGSGRGFLVYVPA"  
4908..7424  
/gene="nbmf"  
/gene="nbmf"  
4908..7424  
/note="beta glucosidase"  
/codon\_start=1  
/transl\_table=11  
/product-"Nbmf"  
/protein\_id="AA08355.1"  
/db\_xref="GI:22001236"  
/translation="YHQAQREKSTVTKTRILPRVRSRTPEPRAFTLAVVGLTAGTIV  
AAAPGAGCTGHVOYTSKAEIVNOMTLEKISVHNALDPPDRNVGLGVPLGIAP  
EIRADGPNIGIRLVGRITAPVAPALASTDDSDMSYGVYMGDRALGQDMVLCR

MANNIRVPHGRNTESEDEPLVSSRTAAVQIKIGIAGILMTTAKHFANNQNNRS  
VNATVDEQTLREIEFPFAEASKAGASFMQAVGVNGKSCQDELLNNVLTQWCF  
OGAVMSDWLPTGTDATITKIDQEMGVELPDGIDPPGSPPAKFFGNGALQVNLNGV  
PEAVTSAERIVNOMDKFGLLATPTAPRPREDRGAQOASRKVAENGAVILNREGA  
LPLADGAKSITAVNGPAPVDPKTYGLTSAHVPPSAAPLDTIKARAGAGVATYEGG  
EETGTGRTPAQOLSPARNQGLPEPKAGALYDCTLVPPADGEFRTVKVANGATATQ  
LGSHTTIEGQYGVGVSSPLKLTIGTRHLITISGRAMATPLSTELGWTTPEADATTA  
KAVESARKARTAIIVAADGTGVDRLSLTPQDGLISAVADANETIYVLTGSS  
VLMPLSKTRAVLDMWYPGQAGAEATNALYGDVNPGLKIOSPPAENQAAVAGDN  
RYPGVDNQOQYSEGIHGVRFWEDEKNKPLPEFGHGLSYTSFTQSAPTVRTSTGGLK  
VTYVYRNSGQARGOEVQVQALYGASPKYTAPOAEKKLVGYTKVALAAGESKTYVNVDR  
KOLQWDAQSDSWRTGIGSRLQTCSSADLKNSQYTV"  
/gene="nbmg"  
/gene="nbmg"  
complement(7514..8653)  
/note="aminotransferase"  
/codon\_start=1  
/transl\_table=11  
/product-"Nbmg"  
/protein\_id="AA08356.1"  
/db\_xref="GI:22001237"  
/translation="MTARTETGRVPEFLDKAAVEELRPETDAIARVLDSGRYLGP  
LEAFABERATYETDHAAGVSGMALQIALRGIGIGDGVIVPSHTYIASMLAVSA  
TGATVPVPEPNDEHPTDLPLVERKAITRTTALPLVHLXGRPADLDALROLADRHGLH  
IVEDAAQAHGARKYRGRIRGASVAAFSFYPEKINLGCFGDGGASTYLAETLSGLGLP  
RNYGSRQYSHETKTSRNLDEIOAAVIRIRLALDMSNGRSTYLAETLSGLGLP  
IALPTAGTDPVWHLFTVYTRDRDELDDHDSRGVDVLTHTVPVHLSPAYAGAPP  
EGSLPRAESFARQVSLPGLPRTDRLTQALRVDAVREWAEOAGR"  
complement(8650..9684)  
/gene="nbmh"  
/gene="nbmh"  
complement(8650..9684)  
/note="NDP-glucose 4,6 dehydratase"  
/codon\_start=1  
/transl\_table=11  
/product-"Nbmh"  
/protein\_id="AA08357.1"  
/db\_xref="GI:22001238"  
/translation="MNLVTGAGGFTGSHFVROLLAGAPVDPAGEVYVLDSTYAGN  
RANLAEVADPRLREYHGBIRADLILARLRCGVDAVHRAASHVDRSITAGSVTGT  
VNOGTOTLQCLADGAGVAVSVNDEYVGSIDGSMNESPJLEFNSPYAASKASDL  
VARAKYRTYGLIYDNDVVDIAKNIRSPRGELITVNVRYLTERGAEVLDJRG  
HTDDHCKGIALYLAGRAEYVHIGGRETNRRELITGLDLSADMSSVRYVADROG  
HDLRSLDGSKELELGYRQIAFDGLAKTTIRWYRENDWMEPLKATAPOPATAPQ  
PATAVGVA"  
complement(9759..10637)  
/gene="nbmi"  
/gene="nbmi"  
complement(9759..10637)  
/note="NDP-glucose synthase"  
/codon\_start=1  
/transl\_table=11  
/product-"Nbmi"  
/protein\_id="AA08358.1"  
/db\_xref="GI:22001239"  
/translation="MKGYLAGGSSGRRLHPATSVISKOILPYNKPMTIYPLSYMLG  
GIREIQIISTPHOITELFOSILGGRHLGIELDAVKEKAGIADALIVYAGHIGDPC  
ALLIGNIHFGPGLVSLDRSIRLADGCVLFGVPVADPEPVAEADATGRLTGLEYK  
PEKPRSHLAVTGLIYDNDVVDIAKNIRSPRGELITVNVRYLTERGAEVLDJRG  
FALDGTGHSLLRAQVYVLEERGVMYIAGIEEIAFMGFIDEACHSISLEGISRT  
EYGSYIEMTAGREGAP"  
complement(10840..12297)  
/gene="nbmj"  
/gene="nbmj"  
complement(10840..12297)  
/note="putative NDP-deoxyhexose 3,4-reductase"  
/codon\_start=1  
/transl\_table=11  
/product-"Nbmj"  
/protein\_id="AA08359.1"  
/db\_xref="GI:22001240"  
/translation="MTAPAPVAVTAPASARCAHPGADIGAAVRAVGOSLAAGCLVPDEA

[illegible]

Query Match	6.28;	Score 184.4;	DB 1;	Length 17665;
Best Local Similarity	48.28;	Pred. No. 4.9e-15;		
Matches 1028;	Conservative	0;	Mismatches 956;	Indels 150; Gaps 12

QY	544	CCCTCGGTTCCACATCTTAACCAAACTCTGGCTCGAAGAGGCGAGTAAGATGATGGCGAAG	603
Db	5290	CCCTGGCCAGACACTTTCACACACTTCATGGCCGACACTTCAGCGAGGGTATGATGGCGCGG	5344
QY	604	AGGCCATCGCTAAGAGTGGCGCATGTGTGATTCCTGGCCCGACTATCAACATGACAAAGCTTCC	663
Db	5350	ACGGACGGCGCGCTGGGCGACAGACATGGTTCTGGGGCCGATGATGAACAAACATCCGGCTGC	5405
QY	664	CTCTGGGTGGAACTGTGGCTTTCGAATGCAATTTGGTAGAATCCGTTCTCTGGCGGGCTTGGGAG	723
Db	5410	CACACGGCGGCGGAACTACGAGACCTTTCAGCAGAGGAGGCCCCCTGGTCTCTCGCGCACCG	5465
QY	724	CTGCGGCTCTCATCCGCGGCAATTTCAGACACTGGAGTGCAGGGCTATGATCAAGCAACTTTT	783
Db	5470	CGGTGCGCCAGATCAAGGGGATCCAGGGGTGGGGTGTGATGACACAGCGCCAAAGCACTTCG	5522
QY	784	TGTGCATATGTCAGAGGACAGCGCGCATGTATGTGTGCAGAGACATGTCACAGAGCGGGCTC	843
Db	5530	CGGCCCAACACAGGAGAGAAACACGGCTTCAGCGTCAACGCGACGGTCGACGAGCAGACGCG	5589
QY	844	TCCGTGAATCTAGACACTCCCGTTCAGATTCCTGTGGAGAGTCCCGACCGCGGGATGGT	903
Db	5590	TCCGGGAATGTGATTTCCCGGCGTTTGAG---GCGTCTCGAAGGCGCGGCGGCTCTCT	5648
QY	904	TCATACGCGGCTACATATGCAATGCGATCANTGGCGGTGTGTGTCAGCGAGAACCCCTAATATCTTG	963
Db	5647	TCATGTGCTCTAATACGGGGGTCAACGGCAAGCGCTCTGGCGCAACGACGAGTGTCTCA	5706
QY	964	ATGGGATCTCTTGAAGGAATGGGGGTGGGATGGCCCTAATCATGAGCACTGTGTACGGCA	10223
Db	5707	ACAAGTCTCTCGCAACGCAAGTGGGGGCTCCAGGGGCTGGGTGATGTCCGATCTGGCTCGGCA	5766
QY	1024	CATACAGTACCAACAGAACCGGTTGTGGCAGGCGCTGCACCTGCAGATGCCCCGACCTCCAC	10838
Db	5767	C---CCCGGAGCGAGCGCATCCACAAAGGGCTCGACACAGAGAT---GGGGCTCAG	5819
QY	1084	GCTTCCGAGGAGAAACACTCAAGTTCAAGCTTCACGGAAGAAAGCCCTTTATCAACGTCA	11433
Db	5820	CTCCCGGCGCAATCCGCGGGCGAGCCCTCCGCGCCGCAAAATGTTCTTGATGATGACGCG	5879
QY	1144	TTGACACAGAGGGCTAAGGAAATTTCTTCAAGTGTGCAGAAAGTGTCTGTCCGCCGAGTGA	1203
Db	5880	CTGA-----ACGAGGCGGTCTTGAGCGGACACGGTCCCGAGGGCG	5919
QY	1204	CGGAAAGCGGCCCGAGACAGCATGTCAACAACACCCGCCAAGAGGGCAGCTCTCTCCGGA	1263
Db	5920	CCGTGACGCGGATCGCGGAGCGCATGTCATCAACAGATGGAACAAGTTGGTCTGTCTCTCG	5979
QY	1264	AGGTTGGCAACGAGGGCATGTGTCTGTCTGAAGACGAGAACACAGTTTGTGCCCTTGAGCA	1323
Db	5980	CGACTCGGGCGCCCGCCCGGACGTCGACAAAGCGGGCGCCCGAGCGGTGTCTCCGCAAGG	6039
QY	1324	AGAAGAAGAAAGACGTGATTTGTGGGCCCGCAAGCGCAAGAGGGCCCATACCAACGGCGGAG	1383
Db	6040	TGCGCGGAAAGCGCGCGGTCTCTCTCGACAGAGAGGGCCAGGCCCTGCGCGCTCCGCGGTG	6099
QY	1384	GCTCTGCCGCACTCAGGGGCTTACAGCAAGTCACTTCCCTTGGACGGCTTCAGCAAGCAGC	1443
Db	6100	ACGCGGCGCAGAGCAATCGCGCATCTGCGCCGACGGCGGTGAGC-CCCAAGTACACGGC	6158

OY	1444	TGAGACGGCGGCATCTGAACACCCTGGCGGCTCACCACCACTTCTCCCATTCCTAAGGG	1507
Db	6159	CTGGGAGAGGCCACAGCTGTCTCCCGGAATCTGGCGGGCGCGCTCGACACATCAAGAAGCC	6218
OY	1504	AGCAGTGCCCTCACGGCCGACGGCGCTCCGGGCAATGCCCCGTGGAGGTTCTTAAGAACCCC	1563
Db	6219	CGGCGGGGGCGGGGTGCGACGGGTGAACGAAAGACGGGTGAGGAGACTTTCCGGGACGGGG	6278
OY	1564	CTGGTACCCCTTACCGGCCACGACATTTGACGAGCTTCTTCCACAAGAGGACATGACC	1622
Db	6279	ATCCCGGGCGGCGACCTAGC-----CCGGCTTCCAMCC	6314
OY	1624	TGTTGAGCTACTTACACACCCCAAGGGGAGACAGCTGTGTAGCCGCAATGTGAAGGCACT	1683
Db	6313	AAGGGCCACCAAGTGGAGGCGGGGGCAAGGGCGGGGGGCTGTACAGACGGCAGCTGACCCTGC	6372
OY	1684	ACACGGCCGACGAGAGACTGTACCTACGAGCTCGGGCTCTGTCTGTGGCGGACGGCAAAG	1743
Db	6373	CCGGCAGACGGCGATACCCGCACTTCGGTCAAAAGGCCACCGGTGGCTACGGCAGGTGAG-	6433
OY	1744	CGTAGGTAGACACAGCTGTGTGCAACAGCCACCAAGCAGTCCCCGGCGATGGCT	1807
Db	6432	-----CTCGGCAATCCACA	6444
OY	1804	TCTTGGECTCGCCACCCGCGAGGAGACGGGCGCATCTCTCTCAAGGGCCAAACAGT	1863
Db	6445	CCATGAGGGCCGCTGAGGTCTACGGCAAGGAGACAGCCCGGCTCTCAAGCTGACCAAGG	6504
OY	1864	ACAAAGTCAAAATGAGTGGTGGGTCCGACCCACCACTACACCCCTCAAGGGCGACACATCG	1922
Db	6505	GCAAGCACAAGT-----CAAGATCTCGGGCTTTCGGAGTGA	6540
OY	1924	TCCCGGCGCAGGCGTCCCTCCGCTGCGCGCTGTCAAGTCAATTGACAGCCAGGCCAAA	1983
Db	6541	GCGCCACCCCTCTCCCTCTGAGGTGGGCTGGGTGGGTGACCGCGGAGCAGCCGACGGCAGCA	6600
OY	1984	TGCAAAATCTCGGTGGCCCTCGCCACAGGAGACGACACCAAGTCAATCTGCGGGGCCCTTA	2043
Db	6601	TGCGCAAGGCGGTGAGTGGCGGCGGGAAGGCCCTGTACGGCGCATCTGTTTTGGG-----	6653
OY	2044	ACGCCGACTGGGAGAACGAGGGGCGCCACGCGCGCGAGATGAAAGTCCCGGCGCTGTGG	2103
Db	6654	--TACGAGCAGCGGACCGAAGGGGCTGCAGCGTCTGGAACCTGTGCGGTGCGCGGTACGGAG	6711
OY	2104	ACCAGCTATTGGCAGAGTGGCGCGCGCGCAACCCAACACCGTGTGTGTATGTAGACAGG	2163
Db	6712	ACAACTGATGTCTGGCGGCTGCGCGACCGCAACCGAATCGATCTGGTGTCTCAACACCG	6771
OY	2164	GCACCCCGGAGGAAATCCCTGGCTGAGACCCACGGCCCGCGATCAACGAGGCGTGTAGG	2223
Db	6772	GTTCGTGCGTCTATGCTGCTGTGCTGTCCAAAGACCCGCGGCTCTGGACATGTGTATCC	6831
OY	2224	GCGGCACGAGACGGGCACTCCATTGGCCGACGCTCTTTGGCGATCAACACCCCTCGG	2283
Db	6832	CGGGCGAAGCGGGCGGAGGCGACCGCGCGCTGTCTACGCTGACGTGAACCCGAGCG	6891
OY	2284	GCAAGCTGTCCCTAGATTCC-----CCAAGCGCTGAGAGACAACCCCGCGTTTCTCA	2337
Db	6892	GCAAGCTACGCAAGAGTTCTCCGCGCGCGGAGAACCAAGACGCCCTGTGCGCGGACCCGA	6951
OY	2338	ACTTCCGACCGAGGCGGGCGACGACT--GTACGGCGAGAGACTCTACGCTCGGTATCA	2394
Db	6952	ACCGCTACCGGGGCTGACCAACCAACAGACGACAGGAGGCAATCCAGTGGGGTAC	7011
OY	2395	GGTACTACGAGTTGGCGACAAAGGAGCTAAATTTCCCTTTGGCAGGGCTGTCTTACA	2454
Db	7012	GCTGTTGCAAGAGAAAGTCAACCGCTTCTCCGTTGGGACGCGCTGTCTGTACA	7071
OY	2455	CCATTTTGGCTTTTCCAAATCTCTCGGTGTCTCAACAAGAC---GGCAAGTGAAGCGGT	2511
Db	7072	CTGTGTTACAGCAGAGAGCGCCGACCGGTGTGTGTGCGACGTCCACGGGCGGCTTGAAGTCA	7131
OY	2512	CCCTCTCTCGTGAAGAACACCGGCTCCGTCGCCGGGCGACAGGTGGCCAG--TTCAGTCA	2571



